THE PREDICTIVE **SOCIAL BRAIN**

ON THE PROCESSING OF OTHER PEOPLE'S BEHAVIOUR

DONDERS
SERIES

Lieke Heil

The predictive social brain:
On the processing of other people's behaviour

Lieke Heil

ISBN: 978-94-6284-156-7

Cover design: Ron Zijlmans
Lay-out: RON Graphic Power, www.ron.nu
Printing: ProefschriftMaken || www.proefschriftmaken.nl

The predictive social brain: On the processing of other people's behaviour

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op

donderdag 13 september 2018
om 12.30 uur precies

door
Lieke Heil
geboren op 5 mei 1987
te Deurne

Promotor
Prof. dr. H. Bekkering

Copromotor
Dr. I.J.E.I. van Rooij

Manuscriptcommissie
Prof. dr. F.P. de Lange
Prof. dr. R.W. Holland
Dr. S. Quadflieg (University of Bristol, Verenigd Koninkrijk)

Contents

1

Introduction

In our daily social interactions, we see other people performing a wide range of actions. Imagine, for instance, watching a friend preparing dinner. You see her cutting vegetables, heating up a pan on the stove and stirring a sauce. In the meantime, you are processing all this information: you might wonder what your friend is doing, what you could do to help and if the dish she is cooking will turn out to taste good or not. Fortunately, your brain will be able to provide all sorts of information to help you understand the situation quickly. It might help you remember what happened the last time she made dinner and activate world knowledge about different flavours, cooking techniques and common ways to prepare food. Therefore, for instance, you might almost instantly realize that she would like you to add some of the salt she is pointing at to the sauce or you might expect a very tasty meal because you know that your friend is a good cook who is using quality ingredients and materials.

In a situation like the one described here, as well as in many other situations, we continuously process information on the behaviour of others in order to understand what they are doing and why they are doing this. Although we usually do this without even realizing, this ability is key to successful social interaction and, ultimately, survival. It not only helps us understand what is happening around us, but it also helps us predict what will happen next and select our own actions accordingly, often even before the other person's action is completed. Although the ease with which we do all this in our daily lives makes it seem so simple, we are only starting to understand how our brains enable the social-cognitive processes involved. Many different sources of information need to be integrated in order to understand why a person performs an action and anticipate his or her next action. This understanding depends, for instance, on what we know about the previous behaviour of that person, the social group they belong to, and the context in which the action is performed. A better understanding of the processes involved would not only improve our general understanding of the brain, but also that of conditions that cause people to experience difficulties in processing other people's actions, such as autism spectrum disorder (ASD) and schizophrenia.

A promising framework for explaining the neural mechanism underlying action understanding and the speed at which we do this is the predictive processing framework. In this thesis, I will explore whether this framework is indeed able to account for different aspects of action processing. To this end, this introduction will discuss what is currently known about the way in which people process and understand other people's actions and how the predictive processing framework may relate to other accounts of social-cognitive processing, before introducing the studies reported in this thesis.

NEURAL MECHANISMS FOR PROCESSING OTHER PEOPLE'S ACTIONS

Different neural mechanisms have been proposed to be involved in action understanding. One very influential proposal involves the mirror neuron system. Mirror neurons are neurons that fire not only when actions are performed, but also when those actions are observed (Rizzolatti & Craighero, 2004), in line with William James' idea that "every mental representation of a movement awakens to some degree the actual movement which is its object" (1890). Support for the existence of this type of neurons was first found in premotor and parietal areas in macaque monkeys (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) and later also in the human supplementary motor area and medial temporal lobe (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010). In addition to this direct evidence for mirror neurons, more indirect support comes from neuroimaging studies that have found areas, typically in inferior parietal and inferior frontal parts of the brain, in which activation is elicited during both performance and observation of actions (Gazzola & Keysers, 2008; Iacoboni et al., 1999; Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Molenberghs, Cunnington, & Mattingley, 2012). Mirroring accounts of action understanding suggest that the mirror neuron system supports action understanding by matching observed behaviour to the motor representations used to perform the same behaviour. In other words, by 'mirroring' the actions of others, we automatically understand them 'from the inside' (Iacoboni et al., 2005; Rizzolatti, Fogassi, & Gallese, 2001; Rizzolatti & Sinigaglia, 2010).

The pattern of activation in another set of brain areas has inspired an alternative account of action understanding. The so-called mentalizing or theory-of-mind network typically includes areas outside of the motor system, such as the medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), posterior cingulate cortex (PCC), and superior temporal sulcus (STS), and seems to be involved in attributing mental states like thoughts, beliefs and desires to other people (Amodio & Frith, 2006; Carrington & Bailey, 2009; Fletcher et al., 1995; Frith & Frith, 2006; Saxe & Kanwisher, 2003). According to mentalizing accounts of action understanding, we understand other people's actions not by simply matching them to our own actions, but by reasoning about the beliefs and mental states of these other people. Support for this account came from studies showing increased neural activity in areas in the mentalizing network when an observed action occurs in an implausible context (Brass, Schmitt, Spengler, & Gergely, 2007), when it is performed with a deceptive intent (Grèzes, Frith, & Passingham, 2004), or when people reflect on its intentionality (De Lange, Spronk, Willems, Toni, & Bekkering, 2008).

Importantly, several attempts have been made to integrate the mirroring and mentalizing accounts of action understanding (De Lange et al., 2008; Keysers & Gazzola, 2007; Kilner, 2011; Uddin, Iacoboni, Lange, & Keenan, 2007; Van Overwalle & Baetens, 2009). These attempts converge on the idea that actions can be understood at different levels and that mirroring processes seem to operate at lower levels than mentalizing processes. For instance, if you see someone giving someone else a gift, you might realize what this person is doing and how this person is doing this (i.e. handing over a gift by moving his hand towards the other person) and at the same time also why the person is doing this (i.e. to make the other person happy). Whereas mirroring processes seem to be involved in understanding the 'how' and potentially the 'what' or immediate goal of an action (Hamilton & Grafton, 2006), mentalizing processes seem to be involved in understanding the 'why' or the intention of the action (Spunt & Lieberman, 2012, 2013). As understanding the intention behind an action requires knowing which action was performed, mirroring processes seem to inform mentalizing processes. This suggests a hierarchical organization in which the immediate goal of an action occupies a lower level than its intention. As such, the complementary accounts of action understanding may be accommodated in the predictive processing framework (Kilner & Frith, 2008), an account that is receiving an increasing amount of attention and promises to reconcile not only social-cognitive but also other types of processing in a unifying theory (Clark, 2013b; Friston, 2005, 2010; Hohwy, 2013).

PREDICTIVE PROCESSING OF OBSERVED ACTIONS

Whereas traditional mirroring accounts tend to describe action understanding as a primarily bottom-up process in which sensory input of observed actions is passively matched to motor representations of actions observers could perform themselves, the predictive processing framework stresses its top-down nature instead. Although this framework was recently proposed as a unifying framework for brain functioning (Clark, 2013b; Friston, 2005, 2010; Hohwy, 2013), it was initially coined to describe visual perception (Lee & Mumford, 2003; Rao & Ballard, 1999). According to hierarchical predictive processing, we do not perceive something because sensory input is transformed into a perceptual interpretation or representation, but rather because this sensory input was correctly predicted by an internal generative model. For example, as you know that your friend is preparing a spicy curry, your generative model may generate a prediction that the red object she just took from a bowl is a chilli pepper. This prediction is sent from high levels that represent rather abstract knowledge about objects and contexts down to more low-level perceptual areas in the brain. If the object is indeed a pepper, no further processing is required and you indeed perceive it as a pepper. If, however, the object is actually an apple, the prediction is not correct and sensory prediction errors arise. They are sent from lower levels to the levels above, in order to adjust (e.g. the value assignments for variables in) the generative models until ultimately, the prediction matches the sensory input.

The predictive processing framework assumes that the brain operates according to Bayesian principles: it represents probabilities and combines priors and sensory evidence in ways that comply with Bayes' theorem (Knill & Pouget, 2004). It further assumes that in case prediction errors arise, they require additional processing in terms of time (e.g. Den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010) and neural activity (e.g. Kok, Jehee, & De Lange, 2012) in order to be 'explained away', as the ultimate aim of predictive processing is to minimize prediction errors (Friston, 2010). In the process of explaining away, the match between the prediction (i.e. the prior) and the sensory evidence is improved. This can be done in several ways, for instance, by revising either the structure or parameters of the generative model or only its value assignments (Kwisthout, Bekkering, & Van Rooij, 2017). In this process, perception may sometimes become biased towards the prediction, as the process of explaining away prediction errors is not aimed at finding the prediction that corresponds to reality, but the prediction that best explains sensory evidence. The framework can therefore explain perceptual biases such as binocular rivalry (Hohwy, Roepstorff, & Friston, 2008) and motion illusions (Weiss, Simoncelli, & Adelson, 2002).

It has been proposed that the framework is able to explain not only visual perception (e.g. Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Kok et al., 2012; Spratling, 2010; Summerfield & Egner, 2009), but also, for instance, auditory perception (e.g. Wacongne et al., 2011), action (e.g. Adams, Shipp, & Friston, 2013) and interoception (e.g. Seth, Suzuki, & Critchley, 2012). Furthermore, it has been suggested that as a general account of brain functioning, it can also be extended to the rather high-level and potentially abstract domain of social cognition (Bach & Schenke, 2017; Kilner, Friston, & Frith, 2007b; Koster-Hale & Saxe, 2013).

In terms of action understanding, this would mean that we use our prior knowledge about actions that we have performed or observed before in order to create a generative model that provides us with predictions that are tested against the observed behaviour. This would lead to faster and more efficient processing of actions that are expected compared to those that are unexpected. Only few studies, however, have investigated if predictive processing indeed provides an explanation for the way in which people process other people's actions. Findings that are in line with this idea are, for example, those showing that top-down and bottom-up signals in the brain are modulated by the probability of an agent-caused event (Van Pelt et al., 2016) and that knowledge about the intention of a reaching or withdrawing action biases the perception of its kinematics (Hudson, Nicholson, Ellis, & Bach, 2016). The aim of this thesis is to further explore whether the framework can explain behavioural and neuroimaging data of studies in which people build up expectations of actions and their outcomes based on, for instance, knowledge about the person performing the action.

GENERATIVE MODELS FOR ACTION PROCESSING

The predictive processing framework assumes that efficient processing depends on useful predictions and it is therefore of key importance to improve our understanding on the generative models in which they arise. Predictions about another person's actions are only informative if they incorporate a wide range of knowledge. For instance, when we meet someone, brain areas involved in representing their physical appearance are linked to areas representing their character traits (Greven, Downing, & Ramsey, 2016) and when we process the action of this person, we consider these characteristics, but also the context in which the action is performed and the objects involved. If we assume that predictive processing provides us with a correct understanding of brain functioning, then all these kinds of knowledge should be represented in our generative models and exert their influence on cognitive processing.

Individual person information in generative models

In order to make meaningful predictions about actions, generative models first need to represent knowledge about individual persons. In the example where you observe your friend preparing a meal, knowing that your friend is an experienced cook who especially likes Indian food will influence your expectation of whether the curry will turn out to taste good or not. Based on, for instance, previous experience with a person, we build up expectations about their character traits and skills. According to predictive processing, violation of these expectations should result in prediction errors that require additional processing in order to be explained away. In line with this idea, previous studies have found that when people read descriptions about another person, neural activity increases when a second description of their behaviour is not in agreement with the trait suggested in the first description (Dungan, Stepanovic, & Young, 2016; Ma et al., 2011; Mende-Siedlecki, Cai, & Todorov, 2012). It is unknown, however, whether knowledge about a person's characteristics also influences the processing of online observation of the actions performed by this person. If this is indeed the case, then the observation of actions or outcomes that are unexpected given knowledge about the person should result in prediction errors, which require additional processing as they are explained away in brain areas involved in social-cognitive processing.

Social group information in generative models

The generative model also needs to be able to represent knowledge about different social groups. If, in our cooking example, you would not observe a friend but someone you have not met before, you would still have certain expectations based on the social group the person belongs to. For instance, you might expect a person from India to cook a better curry than someone from Germany or a person that seems similar to you to cook a dish that you are familiar with. This is because we tend to attribute character traits based on knowledge about social groups, such as stereotypes. These character traits may then guide our predictions about a person's behaviour (Westra, 2017). The idea that information about the social group of a person is integrated in the generative model that provides predictions about that person's actions, is in line with previous studies showing that the way in which we categorize people into social groups influences how we perceive their appearance and actions (for an overview, see Otten, Seth, & Pinto, 2017; Xiao, Coppin, & Van Bavel, 2016) and that stereotype-incongruent information requires additional neural processing (Quadflieg et al., 2011).

There are many ways in which people categorize themselves and others into social groups and one person may be part of many groups at the same time. One very simple and common way to categorize others is by 'ingroup' and 'outgroup'. If people consider someone a member of their our own social group, they tend to evaluate them more positively (Otten & Moskowitz, 2000; Van Bavel & Cunningham, 2009) and act more prosocially towards them (Tajfel, Billig, Bundy, & Flament, 1971). Importantly, this grouping has been found to influence the way in which information about other people is processed. For instance, people are better at recognizing faces of members of their own group (Bernstein, Young, & Hugenberg, 2007) and judge hand movements performed by members of their own group to be faster than those performed by members of the other group (Molenberghs, Halász, Mattingley, Vanman, & Cunnington, 2013). In addition, when people view faces of members of an ingroup the amplitude of the face-sensitive N170 ERP component (Golby, Gabrieli, Chiao, & Eberhardt, 2001) and activation in the fusiform gyrus as well as other brain areas (Ratner & Amodio, 2013; Van Bavel, Packer, & Cunningham, 2008) increases. These findings suggest that whether the observed person is part of one's own or another social group is integrated in the generative model. Although a better understanding could be informative for a range of social psychological and societal issues, it is unknown how the generative models for ingroup members exactly differ from those of outgroup members.

Causal information in generative models

In order to provide useful predictions, the generative model further needs to represent knowledge about the causal structure of the world. It needs to distinguish, for example, between causal and non-causal predictive cues. Causes have a higher predictive value and it therefore seems likely that they play a key role in the generative models underlying our predictions about action outcomes. If you predict the quality of food, the kitchen in which it is prepared might be a useful predictive cue. However, it is not a causal factor, whereas a cue informing you whether the person preparing it is a professional chef or not, on the other hand, is a causal factor. As such a causal factor is usually a better predictor of future events, distinguishing between causal and non-causal factors may therefore be very helpful. Conceivably, this distinction arises because knowledge about causal structures influences what we learn. As we have suggested earlier (Heil, van Pelt, Kwisthout, van

Rooij, & Bekkering, 2014), this would also explain why people only learn to associate a tone with an air puff if they are aware of the link between the two (Clark & Squire, 1998) and why the blocking effect in conditioning depends on by whether people believe that an event is a possible cause or a possible effect of another event (Waldmann, 2000).

Another helpful distinction can be made between causes and enabling factors (Cheng & Novick, 1991). In our cooking example, whether your friend manages to prepare a nice meal will not only depend on her own actions, but also on the ingredients and materials she uses. Even if she is a very good cook, if she uses vegetables that lack taste, the curry will probably not turn out that well. Although your friend's actions cause the outcome (i.e. a prepared meal), the quality of the ingredients and materials is an enabling factor in this situation. In terms of the generative model, this model should thus be able to represent both direct causal and enabling factors. It is unknown, however, how these factors may be represented.

ASSOCIATIVE ACCOUNTS OF ACTION PROCESSING

In addition to predictive accounts, associative and Hebbian accounts have been suggested to provide an explanation of how people process the actions of others. These accounts assume that relatively simple learning rules may underlay the development of mirror neurons (e.g. Giudice, Manera, & Keysers, 2009; Heyes, 2010; Keysers & Gazzola, 2014).

The associative sequence learning (ASL) theory was initially proposed as a theory about the emergence of imitation (Heyes & Ray, 2000) and developed into an explanation of how mirror neurons may have acquired their mirroring properties as a result of sensorimotor experience (Catmur, Walsh, & Heyes, 2007, 2009). In short, it assumes that associations are learned and that whether an association between events is learned depends on their contiguity (i.e. occurrence close together in time) and contingency (i.e. the degree to which one events predicts the other; Cook, Bird, Catmur, Press, & Heyes, 2014) in line with the Rescorla-Wagner model (Cooper, Cook, Dickinson, & Heyes, 2013; Rescorla & Wagner, 1972). On the other hand, Keysers and Gazzola (2014) suggest that this learning depends on Hebbian principles, including temporal precedence, causality and contingency. At a neural level, Keysers and Perrett (2004) suggest that neurons acquire their mirroring properties due to consistent and repeated firing during observation as well as production of actions, which increases the efficiency between pre- and postsynaptic cells in a network of mirroring areas.

Both of these theories focus on how learning from experience can shape the neural systems involved in action understanding. The sensorimotor associations created could then function in either a bottom-up or a top-down fashion. Whereas a top-down type of processing would be in line with the predictive processing framework, other accounts, including classical mirror neuron theories (e.g. Rizzolatti & Sinigaglia, 2010), may rely on a primarily bottom-up type of processing instead.

OUTLINE OF THIS THESIS

In this thesis, I will investigate whether the predictive processing framework is able to explain different social-cognitive processing involved in understanding and predicting other people's actions. More specifically, I will report on a series of studies exploring the potential nature of the generative models underlying predictions, which are also referred to as cognitive or person performance models. In all of these studies, we used a bowling paradigm in which participants viewed animated movies of people playing a bowling game. Based on different types of information, such as knowledge about the player, they could build up predictions of the score and we assumed that violations of these predictions would result in prediction errors. In **chapter 2**, I report on a study in which we investigated whether knowledge about the skill level of an individual player is used in the processing of his actions. More specifically, in this study, we developed a predictive processing model with a hierarchical structure and causal relations between its levels and investigated whether this model was able to explain the pattern of reaction times to questions regarding the player or the score. In **chapter 3**, I will describe an fMRI study in which we tested whether unexpected action outcomes resulted in additional neural activity involved in the explaining away of prediction errors and whether this activation would overlap with activation in brain areas typically associated with social-cognitive processes. In **chapter 4**, I will then report on a study in which we consider whether generative models of another person's task performance differ between ingroup and outgroup whether they distinguish between individuals in both ingroup and outgroup. **Chapter 5** will describe a study focusing on how knowledge about different types of causal factors in the environment may be represented in a generative model. Finally, in **chapter 6**, the findings will be summarized and their theoretical implications will be discussed.

2

One wouldn't expect an expert bowler to hit only two pins: hierarchical predictive processing of agent-caused events

based on: Heil, L., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (2018). One wouldn't expect an expert bowler to hit only two pins: Hierarchical predictive processing of agent-caused events. *Quarterly Journal of Experimental Psychology.* Advanced online publication. doi:10.1177/1747021817752102.

ABSTRACT

Evidence is accumulating that our brains process incoming information using top-down predictions. If lower-level representations are correctly predicted by higher-level representations, this enhances processing. However, if they are incorrectly predicted, additional processing is required at higher levels to 'explain away' prediction errors. Here, we explored the potential nature of the models generating such predictions. More specifically, we investigated whether a predictive processing model with a hierarchical structure and causal relations between its levels is able to account for the processing of agent-caused events. In Experiment 1, participants watched animated movies of 'experienced' and 'novice' bowlers. The results are in line with the idea that prediction errors at a lower level of the hierarchy (i.e. the *outcome* of how many pins fell down) slow down reporting of information at a higher level (i.e. which *agent* was throwing the ball). Experiments 2 and 3 suggest that this effect is specific to situations in which the predictor is causally related to the outcome. Overall, the study supports the idea that a hierarchical predictive processing model can account for the processing of observed action outcomes and that the predictions involved are specific to cases where action outcomes can be predicted based on causal knowledge.

INTRODUCTION

When we observe the actions of another person, we predict what this person is going to do next in order to decide what his or her aim is and to adapt our own response accordingly. To do this, we take the characteristics of this observed person into account. Imagine, for example, that you see a florist reaching for a vase. You could decide to offer help by handing it to him so he can use it to put flowers in. If, on the other hand, you see a small child reaching for the same vase, your prediction of the potential outcome will tell you that this might not be such a good idea, so you can intervene with the child's action and put the vase at an unreachable location. Previous research supports the idea that in order to interact with the world effectively, we use information from previous experiences to predict a specific visual stimulus (Den Ouden, Friston, Daw, McIntosh, & Stephan, 2008; Summerfield & Koechlin, 2008) or the outcomes of other people's actions (Aglioti, Cesari, Romani, & Urgesi, 2008). In the current chapter, we explore the idea that predictions about the outcomes of other people's actions arise in a generative model that has a hierarchical structure and consists of causal relations between different levels of this hierarchy.

This idea has its origin in the predictive processing framework. According to this framework, the brain is continuously predicting the input it will receive (Clark, 2013b; Friston, 2005). This means that our brains process incoming information not in a purely bottom-up fashion, but by a cascade of predictions from higher-level to lower-level representations. The top-down neural signal then consists of predicted states of the world and the bottom-up signal consists of the difference between these predictions and the actual input. These differences are called prediction errors. The framework assumes that if lower-level representations are correctly predicted by higher-level representations this enhances processing, e.g. in terms of speed (O'Reilly et al., 2013). However, when lower-level representations are incorrectly predicted, additional processing at higher levels is required to deal with the prediction errors arising at lower levels.

In recent years, empirical evidence has been found to support the idea that the predictive processing framework successfully describes low-level sensory processing (e.g. Bastos et al., 2012; Kok et al., 2012; Phillips, Blenkmann, Hughes, Bekinschtein, & Rowe, 2015; Rao & Ballard, 1999; for a review see Summerfield & De Lange, 2014). Some researchers have pressed that predictive processing may serve as a general account of brain functioning (e.g. Clark, 2013b; Friston, 2010; Hohwy, 2013) in which case it should also describe processing at higher cognitive levels, including those involved in the processing of agent-caused events. Indeed, it has been suggested that neural responses to biological motion and other people's beliefs and desires are modulated by the predictability of an event, in compliance with the features of predictive processing (Kilner et al., 2007b; Koster-Hale & Saxe, 2013). In accordance with this idea, top-down and bottom-up signals in the brain have been found to be modulated by the probability of an agent-caused event (Van Pelt et al., 2016). Yet, little is known still about the potential cognitive principles that govern the prediction of action outcomes from agent information. In the current chapter, we test whether a hierarchical predictive processing model is, in principle, able to account for the way in which we process the outcomes of other people's actions. We specifically explore a model in which the predictive relations between levels of the generative model are causal.

To study our hypothesis in a naturalistic but controlled setting, we used a behavioural paradigm in which participants viewed animated movies of people playing a bowling game. In these movies, the agent's action (i.e. throwing a bowling ball) caused an outcome (i.e. the pins were knocked down). In Experiment 1, the specific outcome (i.e. the score) could be predicted based on identification of the agent, i.e. knowledge about the performance of this agent on previous trials. There were two agents, indicated as the novice and the experienced player, who usually obtained a low and a high score respectively. In 25% of all trials however, players obtained scores that were incongruent with their skills. According to the model that we explore here, such an incongruence results in a prediction error at the level of the hierarchy at which 'outcome' is represented. In the current bowling set-up, this prediction error should increase processing at the higher level of the hierarchy at which the 'agent' is represented, as this is where the prediction error needs to be 'explained away'. To investigate the idea of a hierarchically organised model, we asked participants to report after each movie either which agent ('experienced' or 'novice') or which outcome ('high score' or 'low score') they observed. For each of these two questions, reaction times were compared for questions following predicted versus unpredicted outcomes. Reaction times have previously been found to correlate with the improbability of an event (Bestmann et al., 2008; Den Ouden et al., 2010) and are therefore assumed to reflect the prediction error. Given the predictive processing model that we investigate here, we hypothesise that the observation of an unpredicted outcome slows down the response to a question about the agent, as the prediction error arising at the outcome level is to be explained away at the agent level, requiring additional processing at that level and thereby slowing down the reporting of the inferred agent information. In other words, we predicted a longer reaction time for the agent question if it followed an unpredicted rather than a predicted outcome, as a prediction error needs to be explained away at this level. On the other hand, as a prediction error is assumed to be explained away at a level above, not at the level at which the error occurs itself, a prediction error occurring specifically at the outcome level is not expected to influence the reaction times for the outcome question. In this way, measuring reaction times for these questions affords us to test the hypothesis of a hierarchically structured predictive model in which information about agents is processed at a higher level in this model than information about outcomes.

Furthermore, we explored the role of knowledge about the causal structure of the world in the prediction of the outcomes of agent-caused events. Predicting what types of outcomes are likely produced by what types of agents presupposes a generative model that encodes such world knowledge. This knowledge tells us that some agents have a propensity to cause some events more often than others. For instance, a skilled bowler is more likely to hit a strike than a novice bowler. Also, we know that agents may cause certain events but not others. A person may cause an object to fall, but not the sun to shine. This seems to be mainly the result of knowledge about the mechanisms that cause events (Shultz, Fisher, Pratt, & Rulf, 1986). The influence of this type of knowledge also explains why young children are more likely to expect the movement of an object to be caused by an agent rather than a by a train (Saxe, Tzelnic, & Carey, 2007), why people only learn to associate a tone with an air puff if they are aware of the link between the two (Clark & Squire, 1998), and why the remembered speed of a movement is influenced by the effect it seems to have caused in Michotte's (1963) launching effect paradigm (Kerzel, Bekkering, Wohlschläger, & Prinz, 2000).

To investigate the hypothesis that a predictive processing model with causal relations between the levels is able to account for the type of processing that occurs at high levels of the hierarchy where agent-caused events are represented, we compare the results of this first experiment to the results of two follow-up experiments. Whereas in Experiment 1 the outcomes of actions (number of pins thrown over by an agent) could be predicted based on knowledge about the agent (his propensity to throw high or low scores), in Experiment 2 we created a situation in which the outcome could only be predicted from a coloured patch next to the agent. Colours, unlike agents, do not cause outcomes. Therefore, if the predictions made in Experiment 1 were specifically based on causal knowledge of agents causing outcomes, none of the effects predicted for Experiment 1 would be predicted for Experiment 2. In Experiment 3, the predictive cue was the agent's shirt colour. Given that according to world knowledge a shirt colour can be a cue to an agent's identity, we would again expect the same reaction time effects in Experiment 3 as predicted for Experiment 1. On the other hand, if the predictions in all experiments would be merely based on non-causal associations between two events, one would expect similar patterns for all experiments. The set of three experiments together allows us to explore whether or not a hierarchical predictive processing model with causal rather than non-causal relations between the levels can account for the processing of other agents' actions. As such, we hope our results will open the way for future research on this topic.

EXPERIMENT 1

Methods

Participants

Twenty-eight healthy, right-handed participants (22 female) between the ages of 18 and 26 (mean age 22.6) were recruited for the first experiment. They were paid 10 euros or received course credits for their participation. The study was approved by the institution's local ethics committee and written informed consent was obtained from each participant.

Stimuli and design

A total of 24 animated movies of a bowling game was created using Autodesk's 3ds Max 2014 and MotionBuilder 2014 (www.autodesk.com). These movies showed a bowling lane and one of two agents, who could be recognised by their clothing. The avatars for the bowling players were selected from Worldviz Vizard Complete Characters (www.worldviz.com/products/avatars/complete-characters). In each movie, the agent threw a ball directed at the pins and disappeared at 1200 ms after the start of the movie, in order to keep the visual display of the action outcome the same for the two agents. The ball then rolled towards the pins, either a little left or right of the centre and hit 1, 2, 3, 6, 7, or 8 pins. The kinematics of the ball movement only varied in terms of the direction and was not associated with a specific agent or outcome. Each movie lasted 5 seconds. By keeping the kinematics of the action constant, we were able to investigate predictions that are based on information about the agent, rather than on kinematics. In 75% of all 288 trials, the outcome was as expected based on the agent's skill level. This means that one agent, who was introduced to the participants as the novice player, received a low score (1, 2, or 3) in 108 out of 144 trials, whereas the other agent, who was introduced to the participants as the experienced player, received a high score (6, 7, or 8) in 108 out of 144 trials. More specifically, within the category of low scores, a score of 2 was most frequent (96 out of 144 trials), as was a score of 7 in the category of high scores (Figure 1). The other outcomes (i.e. 1, 3, 6, and 8) were included as fillers that provided variability in scores and thereby made the experiment more realistic. The movies were presented using Presentation software (version 17.2, www.neurobs.com).

Procedure

Participants were seated comfortably in front of the computer on which the experiment was presented. Instructions were presented on the screen and shortly repeated verbally. In the instructions, it was explained to participants that there were two agents, a novice and an experienced player, who usually (but not always) obtained scores that matched their skill level. In addition, participants were told that after each movie, they would be asked to answer one out of two questions, and that they should pay attention to everything they saw, as they would not know which question would be asked afterwards. They were also instructed to answer the question as quickly as possible. After the instructions, the participants performed four practice trials, in which they received information about which agent they would see, before the movie was presented. This allowed them to associate appearance of the agent with his skill level, as they would need this to perform the task. Immediately after each movie of the actual task, participants were asked to

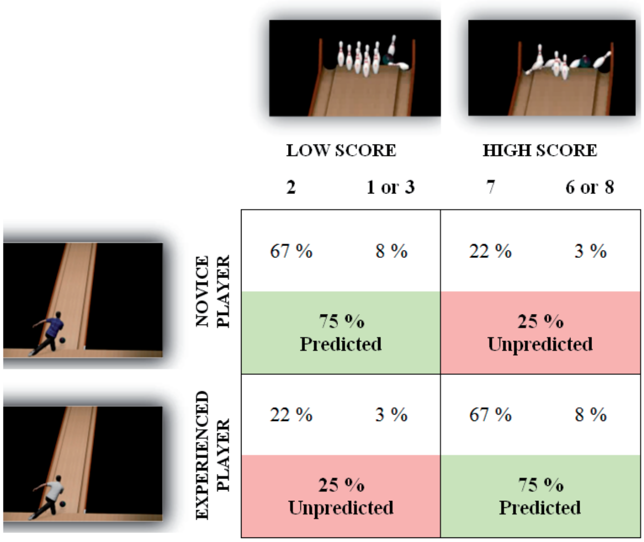


Figure 1. Overview of conditions and stimuli in Experiment 1.

answer one out of two questions, which were presented in random order. One of the questions was about the agent ('Did you see the experienced or the novice player?'), whereas the other question was about the outcome ('Was the score high or low?'). The question was presented on the screen for two seconds, with the two answer options presented underneath. Participants answered by using their index fingers to press either the left button or the right button on a button box as quickly as possible. The order of the answer options was randomised to prevent motor preparation. In the practice trials, participants received feedback on the accuracy of their answer. This was not the case in the actual experiment. After each trial, a fixation cross was presented for a duration that was randomised between 500 and 2500 milliseconds.

Reaction times to the questions that followed movies in which the outcome was 2 or 7 were analysed using a 2 (agent: novice vs. experienced player) × 2 (outcome: 2 vs. 7) × 2 (question: agent vs. outcome) repeated measures ANOVA. The reaction times to questions that were answered incorrectly were not considered in this analysis. As all responses given after 2000 ms were labelled incorrect, our final dataset did not include any trials with very long reaction times. The data were checked for trials very short reaction times (< 100 ms), but no such trials were found. Furthermore, reaction times to questions following movies with outcomes 1, 3, 6, and 8 were also not considered in the analysis, as they were included as fillers in the experiments to make the events look as naturalistic as possible and appeared very infrequently (a total of 32 out of 288 trials). Based on our hypothesis that additional processing is required at the agent level as a result of a prediction error, we anticipated that unexpected events (i.e. the novice agent obtaining a high score or the experienced agent a low score) would bring about longer reaction times than expected events (the novice agent obtaining a low score or the experienced agent a high score). This results in an anticipated interaction effect between agent and outcome, specifically for the agent question.

Results and discussion

The analysis showed a significant 3-way interaction between agent, outcome, and question, $F(1, 27) = 7.41$, $p = .01$, $\eta_p^2 = .22$ (Figure 2). Follow-up analyses showed that for the agent question, there was an interaction between agent and outcome, $F(1, 27) = 18.95$, $p < .001$, $\eta_p^2 = .41$. More specifically, paired-samples t-tests showed that, for this question, unexpected events resulted in higher reaction times for both the novice player, $t(1, 27) = -4.00$, $p < .001$, and the experienced player, $t(1, 27) = 2.60$, $p = .02$. Overall, participants were on average 46.8 ms slower to respond when the outcome was not as would be predicted from the agent's skills, in line with the hypothesis that unexpected events result in a longer reaction time than expected events as a result of additional processing at the agent level. Such an effect did not appear when participants were asked about the outcome, where there was no interaction between agent and outcome, $F(1, 27) = .80$, $p = .38$, $\eta_p^2 = .03$. For this question, there only was a significant effect of outcome on reaction time $F(1, 27) = 19.36$, $p < .001$, $\eta_p^2 = .42$. More specifically, a score of 7 resulted in a reaction time that was on average 36.9 ms longer than the reaction time following a score of 2.

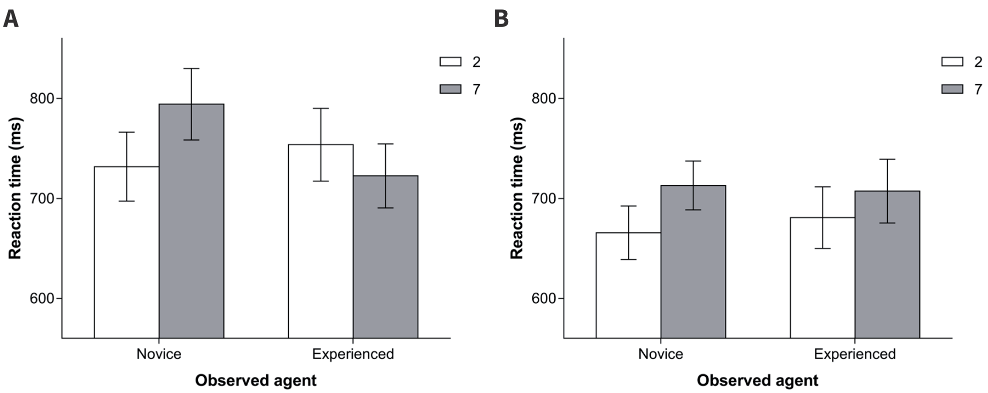


Figure 2. Reaction times (mean ± SEM) for the agent question (A) and the outcome question (B) in Experiment 1, separately for bowler expertise and outcome (score 2 and 7).

Additional analyses confirmed that including the questions following movies that showed another outcome (i.e. 1, 3, 6, and 8) did not influence this pattern of results. Furthermore, the average accuracy over all trials was 95.2% (94.6% for the agent question and 95.8% for the outcome question) and the specific pattern of these data was incompatible with the possibility that the reaction time effects were driven by a speed-accuracy trade off.

The 3-way interaction shows that the prediction effects differ between the two questions. As participants were not aware which question they would be asked after each movie, it is unlikely that they made predictions in one situation, but not in the other. Rather, it seems that specifically for the agent question, reaction times were influenced by the violation of the participant's prediction. This is consistent with the idea that prediction errors at the lower level of the processing of the causal hierarchy, i.e. the outcome level, slow down the reporting of information at a higher level, i.e. the agent level.

In this first experiment, predictions about the outcome of another person's action could be based on a causal relation between the agent and the outcome. To investigate the specificity of this type of prediction to this causal relation, we performed a second experiment in which the score could not be predicted based on the agent's skills, but on an arbitrary statistical relation between a coloured box next to the agent and the outcome. If people's predictions indeed crucially depend on causal knowledge about agents causing outcomes, no effect on reaction times of predictability of events should be observed in Experiment 2.

EXPERIMENT 2

Methods

Participants

Twenty-eight participants (23 female) between the ages of 19 and 29 (mean age 23.1) took part in the second experiment. They were paid 10 euros or received course credits for their participation. The study was approved by the local ethics committee and written informed consent was obtained from each participant.

Stimuli and design

Out of the 24 animated movies from the first experiment, 12 movies with only one of the agents were selected. This means that there were still two ball directions (left and right) and six outcomes (1, 2, 3, 6, 7, and 8). Instead of two different agents, there were now two different coloured boxes (yellow and blue). In each movie, one of these boxes was presented next to the agent. Like the agent, the box was presented from the beginning of the 5-second movie and disappeared after 1200 milliseconds. The colour of the box correlated with the outcome. If one colour was presented the outcome was likely to be low, whereas if the other colour was presented the outcome was likely to be high. The distribution of trials was the same as in the first experiment, with 288 trials in total, of which 75% was as expected based on the colour of the box (Figure 3).

Procedure

The testing procedure was similar to the procedure for the previous experiment. Although the instructions were also largely the same, participants were now told that they would see a blue or yellow box that would indicate if the score is more likely to be high or low. As

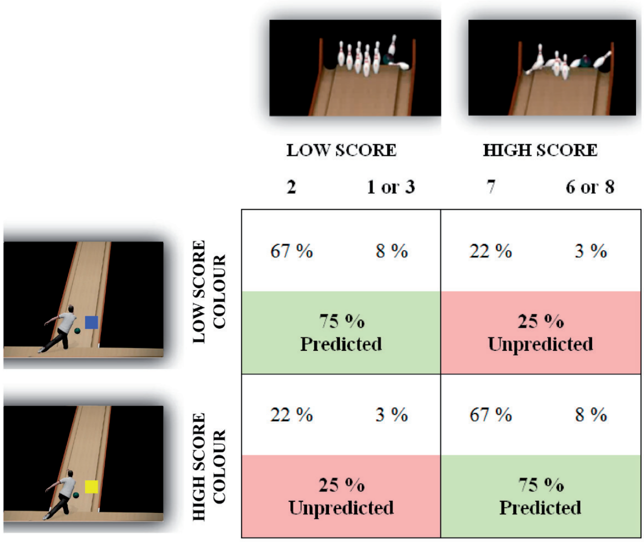


Figure 3. Overview of conditions and stimuli in Experiment 2.

an example, they were told that one of the colours might indicate that the player would probably get a low score, but that this would not always be the case. Again, participants were explicitly informed about the association between the colour and the outcome during the four practice trials. The questions that followed each movie were about the colour ('Was the box blue or yellow?') and the outcome ('Was the score high or low?'). As in the first experiment, reaction times to the questions were measured using a button box and analysed using a 2 (colour: blue vs. yellow) × 2 (outcome: 2 vs. 7) × 2 (question: colour vs. outcome) repeated measures ANOVA.

Results and discussion

Unlike in the previous experiment, no significant 3-way interaction between colour, outcome, and question was found, $F(1, 27) = .27$, $p = .61$, $\eta_p^2 = .01$ (Figure 4). There was also no interaction between colour and outcome, $F(1, 27) = 2.68$, $p = .11$, $\eta_p^2 = .09$. The analysis only showed an interaction between colour and question, $F(1, 27) = 4.33$, $p = .05$, $\eta_p^2 = .14$. To test whether this pattern of results differed from that in Experiment 1, an additional repeated measures ANOVA was run with the same variables as before and experiment as a between-subjects factor. This analysis showed a significant 4-way interaction between colour/agent, outcome, question and experiment, $F(1, 54) = 4.89$, $p = .03$, $\eta_p^2 = .08$, confirming the difference between the experiments.

As in the previous experiment, additional analyses confirmed that including the questions following movies that showed another outcome (i.e. 1, 3, 6, and 8) did not influence this pattern of results. Also, there was no indication that the effects were driven by a speed-accuracy trade off and the average accuracy over all trials was 96.6% (96.2% for the agent question and 97.0% for the outcome question).

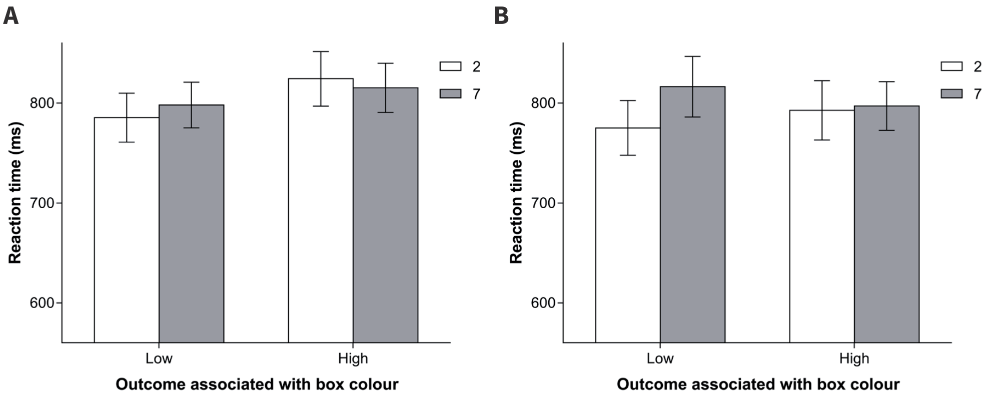


Figure 4. Reaction times (mean ± SEM) for the colour question (A) and the outcome question (B) in Experiment 2, separately for shirt-colour cue (indicative of low or high outcome) and outcome (score 2 and 7).

The results do not provide evidence that participants use the correlation between the coloured box and the outcome to predict the outcome. Whereas in Experiment 1, the correlation between an agent's skill and an outcome had a causal interpretation (i.e. an experienced (or novice) player is more likely to cause an high (or low) score outcome), the correlation between colour and outcome in Experiment 2 did not have a natural causal interpretation (i.e. according to our world knowledge, colours in and of themselves have no causal powers to make pins fall down). The difference between these experiments is in line with the idea that predictions during action observation depend on the causal relation between the predictor and the outcome. However, before concluding that the processing of action outcomes indeed crucially depends on causal knowledge about agents causing outcomes, we need to rule out alternative explanations. In Experiment 1, participants were supposed to infer the agent's identity from his shirt colour in order to answer the agent question. In other words, to distinguish the novice and the experienced agent, participants could not just focus on directly observable cues, as they could for distinguishing the two colours in Experiment 2. To investigate if this difference between the experiments may account for the findings, we conducted a third experiment in which the questions focused only on directly observable information from the movies.

EXPERIMENT 3

Methods

Participants

Thirty-three participants (31 female) between the ages of 19 and 28 (mean age 22.2) took part in this experiment. The sample size was calculated based on the data from Experiment 1, with the assumption that the more implicit causal relation between the colour and the outcome would result in a smaller effect size (it was set at 50% of that of Experiment 1). One participant was excluded from the analyses because the pattern of accuracy for the agent question (i.e. 86.5% correct for the expected outcome versus 15.6% correct for unexpected outcome) suggests that she misunderstood the instructions. As in the previous experiments, participants were paid 10 euros or received course credits for their participation. The study was approved by the local ethics committee and written informed consent was obtained from each participant.

Stimuli and design

This experiment was almost exactly the same as Experiment 1 in terms of stimuli and design. We used the same 24 animated movies and there were 288 trials, with a similar distribution as in the previous experiments.

Procedure

The procedure used in this experiment was similar to the procedure in Experiment 1. The only difference was in the instructions and the questions that were asked after each movie. As we did not want to instruct participants to think about two different agents with different skill levels, we only told them to pay attention to the colour of the agent's shirt without mentioning that there may be different agents. It was explained to them that the colour of the shirt would indicate if the score was more likely to be high or low and thus, that one colour would indicate that the score would probably be low and the other colour that the score would probably be high, although this would not be always the case. Again, the association between the colour and the outcome was mentioned explicitly in the practice trials, but not in the actual experiment. After each movie, participants answered a question about the shirt colour ('Was the shirt purple or white?') or about the outcome ('Was the score high or low?'). Reaction times to the questions were measured using a button box and analysed using a 2 (shirt colour: white vs. purple) × 2 (outcome: low vs. high) × 2 (question: shirt colour vs. outcome) repeated measures ANOVA. Trials were excluded from the analysis in the same way as in the previous experiments. There was one trial in which the reaction time was below 100 ms, but as it was only one, it was not excluded from the analysis.

Results and discussion

Results of the analysis show a 3-way interaction between shirt colour, outcome, and question, $F(1, 31) = 4.21$, $p = .05$, $\eta_p^2 = .12$, as in Experiment 1 (Figure 5). Follow-up analyses showed that for the colour question, there was a significant interaction between colour and outcome, $F(1, 31) = 7.07$, $p = .01$, $\eta_p^2 = .19$. In paired-samples t-tests, unexpected events were found to result in higher reaction times for the colour associated with a low score, $t(1, 31) = -4.28$, $p < .001$, but not for the colour associated with a high score, $t(1, 31) = .17$, $p = .87$. Overall, for the colour question, the reaction time to a question

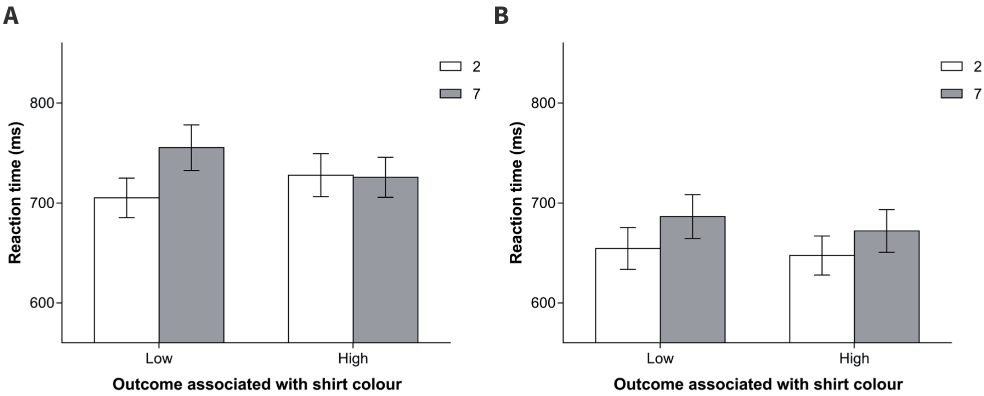


Figure 5. Reaction times (mean ± SEM) for the colour question (A) and the outcome question (B) in Experiment 3, separately for shirt-colour cue (indicative of low or high outcome) and outcome (score 2 and 7).

following an unexpected event was 26.1 ms longer than to a question following an expected event. For the outcome question, there was no interaction between colour and outcome, $F(1, 31) = 0.33$, $p = .57$, $\eta_p^2 = .01$. For this question, there was a significant effect of outcome on reaction time $F(1, 31) = 12.29$, $p < .01$, $\eta_p^2 = .28$. A score of 7 resulted in a reaction time that was on average 28.2 ms longer than the reaction time following a score of 2. This pattern of results resembles that of Experiment 1. A repeated measures ANOVA with the variables from the previous analysis as well as experiment (1 vs. 3) as a between-subjects factor indeed showed no 4-way interaction between agent/colour, outcome, question, and experiment, $F(1, 58) = 0.67$, $p = .42$, $\eta_p^2 = .01$.

Again, additional analyses confirmed that including the questions following movies that showed another outcome (i.e. 1, 3, 6, and 8) did not influence this pattern of results. As in the other experiments, there was no indication that the effects were driven by a speed-accuracy trade off and the average accuracy over all trials was 96.5% (95.2% for the agent question and 97.7% for the outcome question).

So even though participants were not explicitly instructed about different agents with different skill levels, the pattern of results for this experiment is very similar to that of the first experiment. This suggests that the difference between the first two experiments was not simply caused by the fact that one of the questions in Experiment 1 was about the agent, which was not directly observable, whereas in Experiment 2, both questions were about directly observable factors. In Experiment 3, like in Experiment 2, the questions focus on colour and outcome, both of which are directly observable. The pattern of results, however, resembles that of Experiment 1. This is in line with the idea that in both Experiments 1 and 3, colour is used as a cue to the agent's identity and the outcome is then predicted based on this identity. According to this idea, a causal relation between predictor (agent) and predicted (outcome) is crucial for predicting the outcomes of other people's actions.

To test the idea that predictions enhance processing in terms of speed, we ran an additional analysis in which we compared the overall reaction times for all three experiments, using a one-way ANOVA. This analysis indicated that the average reaction times differed significantly between the experiments, $F(2, 85) = 5.16$, $p < .01$, $\eta_p^2 = .11$. Bonferroni-corrected post hoc comparisons revealed that the average reaction time in Experiment 2 (M = 798.8 ms, SD = 124.57) was significantly higher than in Experiment 3 (M = 693.81 ms, SD = 108.87), $p < .01$, whereas the average reaction times for Experiment 2 were higher than for Experiment 1 (M = 715.64 ms, SD = 159.52), although this difference was only marginally significant, $p = .06$. For Experiment 1 and 3, reaction times did not differ significantly, $p = 1$.

Although the questions answered in the three experiments were different, the length and difficulty of the questions in Experiment 2 cannot account for these findings. Rather, the differences in reaction times between the experiments seem to suggest that predictions that participants made in Experiment 1 and 3 allowed them to respond quickly, whereas the inability of participants to use the relation between colour and outcome to predict the outcome in Experiment 2 resulted in a longer reaction time. This in line with the idea that predictions speed up cognitive processing.

DISCUSSION

In a series of experiments, we investigated whether a predictive processing model is, in principle, able to account for the way in which we process the outcomes of other people's actions. More specifically, we explored a model in which predictions arise in a generative model that has a hierarchical structure and consists of causal relations between different levels of this hierarchy. The present results support the idea that such a model can account for this type of processing. To further improve the interpretation of the data, we developed a computational characterisation of this hierarchical predictive processing model in order to assess to what extent the present experiments' qualitative pattern of results is consistent with our theoretical assumptions. A detailed formal description of this characterisation and the associated processes can be found in the appendix. Here, we will briefly explain the characterisation (for a simplified version, see Figure 6), before outlining the relation with the experimental findings.

Crucially, the model consists of three levels (agent, outcome and visual input), hierarchically ordered from top to bottom. Predictions are sent from higher to lower levels. This means that the visual input (i.e. the actual configuration of pins falling down) is ultimately predicted based on the observed shirt colour, through predictions at the outcome and agent levels. The relations between the three levels are causal in nature (black arrows), whereas the relationship between colour and agent (dashed arrow) is deterministic. This deterministic relation indicates that the colour of the agent's shirt

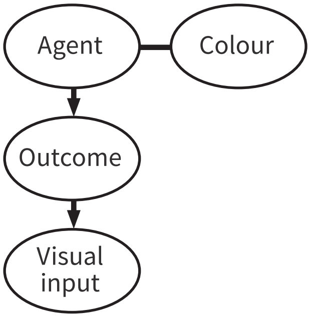


Figure 6. A simplified version of our precise characterisation of hierarchical predictive processing.

can be used to identify the agent. This identification then drives a prediction for the most probable outcome (two versus seven pins), which in turn drives predictions about which exact pins will fall down.

Bottom-up input at the lowest level (i.e. a visual stimulus) that is not fully predicted generates prediction errors. These prediction errors represent information about the input that was not already anticipated (Rao & Ballard, 1999) and are instrumental in updating the hypothesis (explaining away the prediction error) at a higher level. In the predictive processing framework, these prediction errors are weighted, which means that top-down and bottom-up information is balanced by altering the gain on specific prediction error units (Clark, 2013a; Friston & Kiebel, 2009). The prediction error that arises at the level of visual input, and is processed at the outcome level above it, carries a very low weight because of the irreducible uncertainty in this information. This irreducible uncertainty stems from the fact that each outcome is inherently associated with many potential configurations of visual input (e.g. there are many ways in which two out of ten pins can fall down). As a result of this low weight, little additional processing is going on at the outcome level even if some prediction errors arise at the visual level below. On the other hand, the prediction error that is sent to the agent level carries more weight, as the relation between agent and outcome has much more reducible uncertainty and thus allows for updating. Therefore, the discrepancy between the predicted and the perceived outcome (i.e. the number of pins hit) will lead to additional processing at the agent level to explain away the high prediction error. Presumably, this takes time and as a consequence, when participants answer a question that calls for information from this level, the reaction time increases monotonically with the size of the prediction error at the level below. In this case, the increase in reaction time reflects 'explaining away' of the prediction errors by updating the hypothesis about the current situation. More informally, we could say that when participants observe an experienced player hitting only a few pins, their prediction that the outcome is most probably high is violated, leading to a relatively high prediction error, but this prediction error can be explained away based on the knowledge that unpredicted outcomes occur every now and then.

After the prediction error is explained away, the information represented at this level can be read and the question can be answered.

With this model in mind, we will now take a closer look at the experimental results. In Experiment 1, we found that prediction errors at the lower (outcome) level indeed slowed down reporting of information at the higher (agent) level. This is consistent with the idea that the generative model created to predict human actions is structured hierarchically, as we assume that prediction errors are sent upwards from lower levels to higher levels, increasing reaction times at this higher level as they are explained away. Given that only responses to the agent question and not the outcome question were influenced, a non-hierarchical representation of the generative model now seems less plausible.

In two follow-up experiments we investigated the idea that, at high levels of the hierarchy where agent-caused events are represented, the predictive relations between levels of the generative model have a causal representation. In Experiment 1, participants could use their knowledge about the agent to predict the outcome of a bowling action. A novice player was expected to obtain a low score, whereas an experienced player was expected to obtain a high score. The relation that is learned here could be represented as a causal one: based on general knowledge about the world, we know that agents cause outcomes. So when participants learned which agent usually got which scores in the bowling game, they could use this knowledge to base their prediction about the outcome on. In case this knowledge about agents and the causal structure of the world would not be important, one would expect that changing the agent information to another informative cue would result in the same pattern of reaction times. Therefore, in the second experiment, we used a cue with the same predictive probabilities as the agent. It has previously been suggested that predictions are based on simple associations and the contiguity and contingency between a cue and an outcome determine whether a prediction will be made (Keysers & Gazzola, 2014). Contiguity describes the paired occurrence of the cue and the outcome, whereas contingency implies that one event (i.e. the colour cue or the agent) reliably predicts the other (i.e. the outcome). Thus, the contiguity and contingency between the cue and the outcome were exactly the same in all three experiments. The second experiment differed from the other experiments in terms of the type of relation that could be learned. Here, this relation was not a causal one between an agent and an outcome, but a purely stochastic one between a colour and an outcome. In terms of the model presented in Figure 6, predictions about action outcomes cannot be based on a colour since this colour is not intrinsically linked to an agent, as the representation of colour does not have a causal relation with the representation of the outcome at the level below. This is in agreement with a suggestion that we have made elsewhere (Heil et al., 2014): the degree to which two events are associated can be guided by beliefs about the causal structure of the world. For example, people can only be conditioned to blink their eyes when they hear a tone in case they are aware of the relation between the tone and a puff of air to the eye that caused them to blink (Clark & Squire, 1998). In the same way, Waldmann (2000) found that the blocking effect in conditioning (i.e. the effect that the association of an event A with event Y is prevented if A is presented together with another event B that has previously been associated with event Y) is modulated by whether the participants were led to believe that A and B were either possible causes or possible effects of Y. It seems that awareness of a causal relation between two events influences the way in which these events are processed: if people do not interpret two events as causally related, they are not as easily associated, as is the case when an arbitrary colour correlates with a certain outcome in our bowling setting. We therefore suggest that the lack of a causal interpretation in Experiment 2 results in a pattern of reaction times that does not distinguish between predicted and unpredicted outcomes, even though participants were made aware that there was a statistical relation between the colour of the box and the outcome.

In Experiment 3, participants could again base their predictions on a causal relation between an agent and an outcome, but this experiment was designed to control for some of the differences between the two previous experiments. We used the same stimuli as in the first experiment, but asked the participants to answer questions about directly observable features, as in the second experiment. Next to reporting the outcome, participants reported the colour of the agent's shirt. Similar to Experiment 1, reaction times to the colour question (which is comparable to the agent question) were again influenced by predictions about the outcome. These findings are in line with the idea that the brain's predictive model of other agent's actions consists of causal relationships between different levels in a hierarchy.

Previous studies investigating the role of causality in cognitive processing already showed that whether or not two events are perceived to be causally related depends on prior knowledge. This knowledge may, for example, concern the causal mechanism (Shultz et al., 1986), the probability that a causal relation exists or the assumed functional form (i.e. deterministic or probabilistic) of this relation (Griffiths, Sobel, Tenenbaum, & Gopnik, 2011). Importantly, the current study extends these findings by showing that if events are perceived to be causally related, this allows for the prediction of one event based on the other, which then enhances cognitive processing. It seems that the world model that we use to predict other people's actions and their outcomes revolves around causes and their effects.

This does not mean that it is impossible to make predictions based on arbitrary cues. Previous studies actually showed prediction effects for arbitrarily associated events (e.g. Kok et al., 2012). Conceivably, in the clearly non-arbitrary setting that we created in our experiment relevant world knowledge is activated, whereas this is not the case for studies in which arbitrary events are associated. For example, in the study by Kok et al. (2012), an auditory cue was the only source of information available to predict the orientation of the visual stimulus. The richness of information in our experiments may have made it more difficult to associate an arbitrary cue with an action outcome. This potential difference in the associations that are learned in arbitrary versus non-arbitrary settings demonstrates the importance of naturalistic, ecologically valid experimental designs in which a causal relation can be inferred. For example when studying hierarchical predictive processing of agent-caused events, as we do here, it is important to take into account that the structure of the hierarchy may depend on the causal relation between agent and outcome. We suggest that knowing that agents cause outcomes allows us to predict an outcome based on knowledge about the agent.

Importantly, predictive processing is assumed to enhance processing in terms of speed. A comparison between the overall reaction times in the different experiments indeed revealed that in cases in which we found that participants were able to predict the outcomes (i.e. Experiments 1 and 3), processing was speeded up compared to cases in which we found no evidence for these predictions (i.e. Experiment 2). This is in line with the idea that predictions enhance processing in terms of speed and shows the potential benefit of predictive processing as a more general mechanism.

Although our results show that the predictive processing framework is able to account for reaction time effects in the processing of another person's action outcomes, it is also important to consider possible alternative explanations. For instance, in contrast to predictive processing, other accounts may not assign a key role to predictions and prediction errors. An example of such a non-predictive explanation would be an account in which probabilistic inference takes place only after the observation of the events. One could, for instance, hypothesise that the inference of a less probable event requires more processing time than inference of a more probable event, resulting in a longer reaction time for an agent that is not probable given the outcome. Such an account can be substantiated, however, only insofar as there is a plausible mechanism that explains why lower probability events take longer to process. Even if this is granted, such an account would need to make additional assumptions to explain the exact pattern of results. For example, without additional assumptions, the account cannot explain the difference between the two questions, as both need to be inferred and both deal with a similar pattern of low and high probability.

However, it is important to note that many models of cognition may actually be integrated in the predictive processing framework. For example, claims arising from associative theories, such as the associative sequence learning model, do not necessarily oppose those of the predictive processing framework. Press, Heyes, and Kilner (2011) argue that the associative sequence learning model explains how relations are learned, whereas predictive processing explains how learned relations support inferences about other people's actions. In this sense, the two models complement each other since they simply address different questions.

Future research is needed to investigate if a predictive processing model with a hierarchical structure and causal relations between its levels can also account for different types of cognitive processing. Also, it would be interesting to further disentangle the different processes that may underlie explaining away of the prediction error. In case of hypothesis updating, the probability distribution over the candidate hypotheses is reassessed while the generative model remains stable. On the other hand, revision of the generative model, sensory sampling to obtain more information on the unpredicted event, and active inference have been proposed as alternative ways to bring prediction and sensory input closer together (Kwisthout et al., 2017). For example, if the relation between a certain agent and a certain outcome changes in the course of an experiment (e.g. when the novice bowler gets better), participants need to revise their model.

APPENDIX

In this section, we present a precise computational characterisation of predictive processing that makes insightful how the data can be interpreted. It allows us to assess to what extent the qualitative pattern of results of the described experiment is consistent with our assumptions about hierarchical predictive processing. The computational characterization has several components: an information structure and a set of computational processes that operate on this information structure, in line with the proposal in Kwisthout et al. (2017). Below we will describe both the *general* computational model and the task-*specific* instantiation of this computational model.

General information structure

In the predictive processing account, the brain continuously predicts its inputs in a cascade of probabilistic predictions from higher-level cognitive representations to lower-level sensory inputs. The information structure that we use in our computational characterisation of this account is a *set of causal Bayesian networks* B_{ij}. Each network B_{ij} describes a generative model, i.e. a probabilistic relation between possible causes (at level *i*) and predicted consequences of these causes (at level *j*). In our computational characterisation, the causal Bayesian networks consist of a set of hypothesis nodes (denoted by *Hyp*; jointly representing a set of possible causes), a set of prediction nodes (denoted by *Pred*; jointly representing a set of possible consequences), and a set of other nodes (denoted by *Int*; representing contextual influences and modulating complex dependences between *Hyp* and *Pred*) (Figure A1). In the hierarchy, the prediction nodes *Pred* in network B_{ij}, representing level *j* of the hierarchy, are identified with the hypothesis nodes *Hyp* that drive predictions (at level *k*) in network B_{jk}. The network further consists of a set of (discrete) conditional probability distributions, representing the strengths of the stochastic relationships in the network.

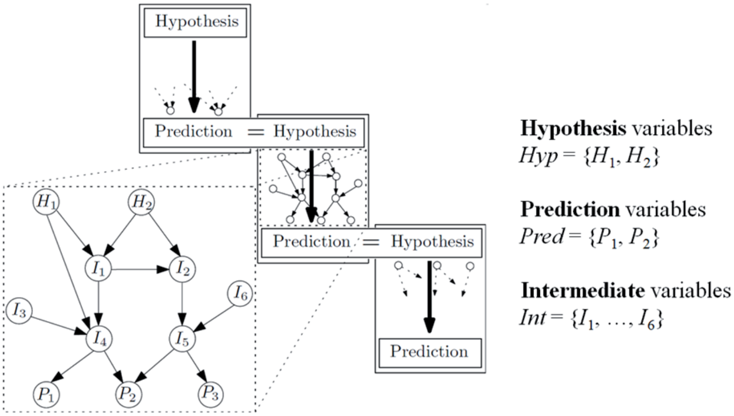


Figure A1. Illustration of the general information structure where the stochastic dependencies between two levels are mediated by a Bayesian network, consisting of hypothesis variables (the set Hyp), prediction variables (the set Pred), and intermediate variables that represent contextual effects or that model complex structural dependencies between hypotheses and predictions.

The arcs in the network can have a causal interpretation, that is, an arc from node *A* to node *B* represents that *A* is a possible cause of *B*. The hypothesis nodes are source nodes (have no incoming arcs *within one level of the hierarchy*) and the prediction nodes are sink nodes (have no outgoing arcs *within one level of the hierarchy*).

Task-specific information structure

The task-specific information structure that we use to characterise the experimental setting of Experiments 1 and 3 represents a hierarchy with three levels: Agent (level 1), Outcome (level 2), and Visual Input (level 3) as depicted in Figure A2. The structure thus consists of two networks B_{12} and B_{23} and contains the following nodes: B_{12} = {*Hyp* = {Agent}, *Pred* = {Outcome}, *Int* = {Colour}} and B_{23} = {*Hyp* = {Outcome}, *Pred* = {Visual input}, *Int* = ∅}. With these networks we associate the following conditional probability distributions: {P(Colour), P(Colour | Agent), and P(Outcome | Agent)} and {P(Outcome), P(Visual input | Outcome)}, respectively. The variables Agent, Outcome, and Colour are all binary, with Agent taking the values ‘experienced’ and ‘novice’, Outcome taking the values ‘2’ and ‘7’, and Colour taking the values ‘white’ and ‘purple’. The variable ‘Visual input’ takes as values every possible combination of any subset of the ten pins, that is, {∅, {1}, {2}, …, {1,2}, {1,3}, …, {1,2,3,4,5,6,7,8,9,10}}. The cardinality of this variable is $1 + \binom{10}{9} + \binom{10}{8} + \ldots + \binom{10}{1} + 1 = 1024$. The conditional probabilities are as follows: P(Colour) is uniformly distributed, P(Colour | Agent) is deterministic, P(Outcome = 2 | Agent = novice) = 0.75, P(Outcome = 7 | Agent = novice) = 0.25, P(Outcome = 2 | Agent = experienced) = 0.25, and P(Outcome = 7 | Agent = experienced) = 0.75. The conditional probability distributions P(Visual input | Outcome = 2), respectively P(Visual input | Outcome = 7), are such that a combination

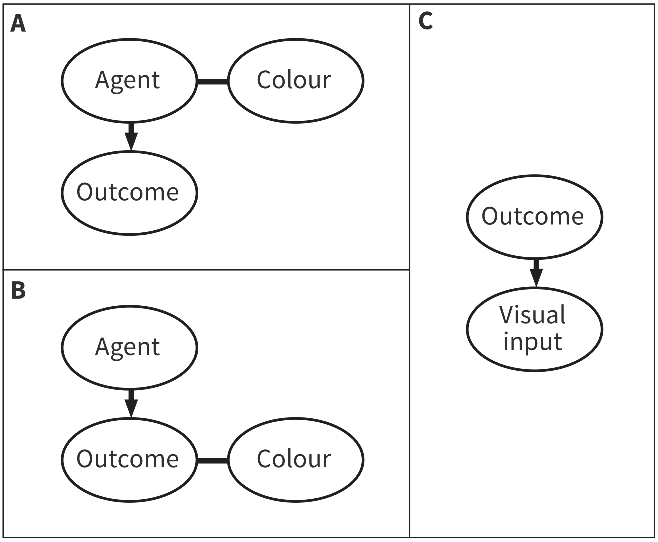


Figure A2. The Bayesian networks connecting the Agent and Outcome levels (top left panel A for Experiments 1 and 3, with Colour as contextual variable influencing Agent; bottom left panel B for Experiment 2, with Colour as contextual variable influencing Outcome) and the Outcome and Visual Input levels (right panel C). The conditional probability distributions in the left network are given in the figure. The conditional probability distributions associated with these networks are given in the text.

of other than two pins (respectively seven pins) falling over has a probability of zero and the probability of any combinations of two pins (respectively seven pins) is uniformly distributed over the $\binom{10}{2} = 45$ (respectively $\binom{10}{7} = 120$) possible combinations.

In Experiment 2, we disassociate 'colour' and 'agent'. This is reflected in our model by the conditioning of Outcome on Colour, rather than Agent on Colour. The causal relation between Agent and Outcome is still in place – as outcomes are caused by agents – but as there are no ways for participants to recognise the expert and the novice player, the statistical dependence of the various outcomes on the agent is de facto absent. We assume in our model that the absence of the causal relation between Agent and Outcome makes it much harder to learn the statistical regularities in the experiment. There is no *a priori* (on the basis of world knowledge) causal relation between Colour and Outcome; there is only an artificially created experimental association. This requires the participants to 'unlearn' the expectation that there is no effect (leading to a uniform prediction of Outcome given Colour) in the course of the experiment, and learn the statistical association in the experiment. This is in contrast with Experiments 1 and 3 where there *is* a causal relation between Agent and Outcome, where the shirt's colour is an indication of the Agent and thus relevant for determining the expected Outcome.

Computational processes (general)

On the information structure described above, we define the following computational processes: **Prediction** (computing the prediction, i.e. the posterior probability distribution for the *Pred* nodes), **ErrorComputation** (computing the prediction error, i.e. the divergence between the predicted and observed distribution for the *Pred* nodes), and **ExplainingAway** (minimizing prediction error, i.e. computing an alternative prior probability distribution over the *Hyp* nodes that minimises the divergence between predicted and observed distribution for the *Pred* nodes). We now formally explicate these computational processes as follows:

Prediction

Input: A causal Bayesian network *B* as specified above.
Output: The posterior probability distribution P(*Pred*) over the prediction nodes *Pred*.

ErrorComputation

Input: A predicted probability distribution P(*Pred*) and an observed probability distribution P'(*Pred*) over a set of prediction nodes *Pred*.
Output: The prediction error δ(P'(*Pred*), P(*Pred*)), defined as the result of a component-wise subtraction of the two probability distributions, i.e. $\forall i$ P'(*Pred* = *i*) – P(*Pred* = *i*).

Explaining Away

Input: A causal Bayesian network *B* as specified above, including a prior probability distribution P(*Hyp*) over the hypothesis nodes *Hyp* and a prediction error δ(P'(*Pred*), P(*Pred*)).
Output[1]: A revised prior probability distribution P'(*Hyp*) = P(*Hyp* | *Pred*) × P'(*Pred*).

In addition, we need to define the following concepts. The *size* of the prediction error δ(P'(*Pred*), P(*Pred*)) is the Kullback-Leibler-divergence D_{KL}(P'(*Pred*) || P(*Pred*)) = Σi P'(*Pred* = *i*) × $\log_2$ (P'(*Pred* = *i*) / P(*Pred* = *i*)). We define the *precision* of the prediction error as *w* × D_{KL}(P'(*Pred*) || P(*Pred*)), where *w* is defined as the *weight* of the prediction error (Clark, 2013). It represents a measure on the amount of *reducible* uncertainty in *B*; if there is a high confidence that *B* is a faithful representation of the actual stochastic dependences between *Hyp* and *Pred*, then the prediction error carries only information regarding the *irreducible* uncertainty that is inherent in the environment.

1 Note that the predictive processing account assumes that only the prediction error δ(P'(*Pred*), P(*Pred*)) [and *not* P'(*Pred*) itself] is processed in the process of explaining away prediction errors. However, as we can always recover P'(*Pred*) from *B* and δ(P'(*Pred*), P(*Pred*)), the given input-output mapping is not underdefined.

Computational processes (task–specific)

The computational processes defined above are concretised as follows on the task-specific information structure:

Prediction

Input: B_{12}, respectively B_{23}, as described above.

Output: For B_{12}, the posterior probability distribution P(Outcome) = P(Outcome | Agent) × P(Agent | Colour) × P(Colour); for B_{23}, the posterior probability distribution P(Visual input) = P(Visual input | Outcome) × P(Outcome).

ErrorComputation

Input: P(Outcome) and P'(Outcome) = P'(Outcome | Visual input) × P'(Visual input), respectively P(Visual input) and P'(Visual input)).

Output: δ(P'(Outcome), P(Outcome)), respectively δ(P'(Visual input), P(Visual input)).

Explaining Away

Input: B_{12}, respectively B_{23}, as described above; δ(P'(Outcome), P(Outcome)), respectively δ(P'(Visual input), P(Visual input)).

Output: P'(Agent) = P(Agent | Outcome) × P'(Outcome), respectively P'(Outcome) = P(Outcome | Visual input) × P'(Visual input).

The weights w_{12}, respectively w_{23}, of the prediction error δ(P'(Outcome), P(Outcome)), respectively δ(P'(Visual input), P(Visual input)), represent the confidence in B_{12}, respectively B_{23}. We make the assumption that the confidence in the generative model that relates agents to outcomes is *less* than the confidence in the generative model that relates outcomes to visual stimuli: the agent-outcome relationship is learned over the course of the experiment, whereas participants have a lifetime experience with the conceptual interpretation 'two pins fall down' of a particular visual stimulus where two pins fall down. Prediction errors in B_{23} represent irreducible uncertainty, i.e. uncertainty that is inherent to the stochastic relation between outcomes and visual stimuli. Therefore, the prediction errors should not lead to revision of the generative model. Given that there is much more reducible uncertainty (and thus a lower precision) in the prediction error at the B_{12} level, the weight of the prediction error is higher at the B_{12} level than at the B_{23} level. We therefore assume that w_{12} is much higher than w_{23} and that the *weighted* prediction errors are lower in B_{23} than they are in B_{12}. As the reducible uncertainty in the prediction error at the outcome level in this experiment is even higher than in Experiments 1 and 3, we assume that the weight of the prediction errors w_{12} is considerably higher in Experiment 2.

Example calculation

On observation of the colour of the player's shirt (purple) the participant expects the experienced player. For this player, the predicted outcome will be P(Outcome = 7) = P(Outcome = 7 | Agent = experienced) × P(Agent = experienced | Colour = purple) × P(Colour = purple) + P(Outcome = 7 | Agent = novice) × P(Agent = novice | Colour = purple) × P(Colour = purple) = 0.75 × 1 × 1 + 0.25 × 0 × 0 = 0.75. Likewise, P(Outcome = 2) = 0.25 × 1 × 1 + 0.75 × 0 × 0 = 0.25. Assuming a uniform probability over all visual stimuli that match the outcome, the prediction for the visual input P(Visual input) = P(Visual input | Outcome = 7) × P(Outcome = 7) + P(Visual input | Outcome = 2) × P(Outcome = 2). Given that P(Outcome = 7) = 0.75 and that P(Outcome = 2) = 0.25, this will generate a probability of 0.0063 for each of the 120 combinations of 7 pins, and 0.0056 for each of the 45 combinations of 2 pins. In contrast, when the white shirt is observed, giving an expectation for the novice player, the predicted outcome would be P(Outcome = 7) = 0.25 and P(Outcome = 2) = 0.75, and P(Visual input) would be 0.0021 for each of the 120 combinations of 7 pins, and 0.0167 for each of the 45 combinations of 2 pins. On observing the actual visual input (that is, a particular combination of 2 or 7 pins falling down), the KL divergence is computed to be 5.91, 8.91, 7.49, and 7.32 for the conditions {novice agent, outcome 2; novice agent, outcome 7; experienced agent, outcome 2; experienced agent, outcome 7}. This means that the prediction error is higher for situations in which the outcome is 7 as compared to those in which the outcome is 2, as the number of possible combinations in which the pins can fall down is higher for an outcome of 7 than for an outcome of 2. For parsimony, we assume the simplest mathematical relationship between the prediction error and reaction time: the processing time of explaining away prediction errors (as indexed by the reaction time) increases monotonically, but not necessarily linearly, with the size of the prediction error. Then, the expected qualitative pattern of the reaction times for these four experimental conditions is consistent with the weighted prediction errors at the B_{12} level as depicted in Figure A3, left panel. The inferred outcome (either P(Outcome = 2) = 1 or P(Outcome = 7) = 1, depending on the actual visual stimulus) is compared with the predicted outcome (either P(Outcome = 2) = 0.25 and P(Outcome = 7) = 0.75, or P(Outcome = 2) = 0.75 and P(Outcome = 7) = 0.25, depending on whether the experienced or the novice agent is observed). This gives a KL divergence of 0.425, 2, 2, and 0.425 for the conditions {novice agent, outcome 2; novice agent, outcome 7; experienced agent, outcome 2; experienced agent, outcome 7}. The expected qualitative pattern of reaction times for these four experimental conditions is consistent with the weighted prediction errors at the B_{23} level[2], as is depicted in Figure A3, right panel.

2 Note that other aspects than weighted prediction errors also influence the reaction times. For example, as it takes longer for seven pins to fall down compared to two pins, it also takes longer for the participant to infer the actual visual input.

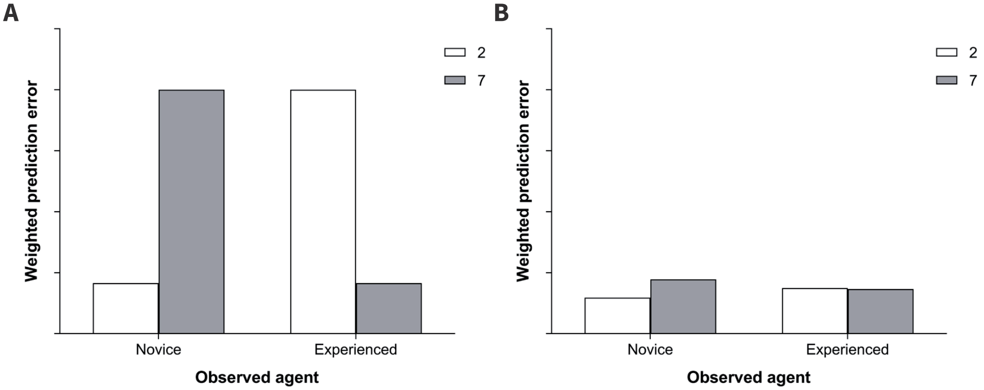


Figure A3. The expected reaction time (on a unit scale) for the four experimental conditions in Experiments 1 and 3, based on the weighted prediction errors at B_{12} (A) and at B_{23} (B).

In Figure A4 the expected pattern of the reaction times is given, based on the weighted prediction errors that follow from the model describing Experiment 2. As described above, we anticipate that there will be a much weaker association between box colour and outcome and a higher weight of the prediction error. Both are due to the artificial (experimental) relation between box colour and outcome that forces the participants to *unlearn* their expectation that there is no effect of box colour. The main consequence thereof is that the weighted prediction errors will be overall higher than in Experiments 1 and 3, and that the effect of colour on outcome will be much weaker than the effect of agent (or colour) on outcome in Experiment 1 (or 3).

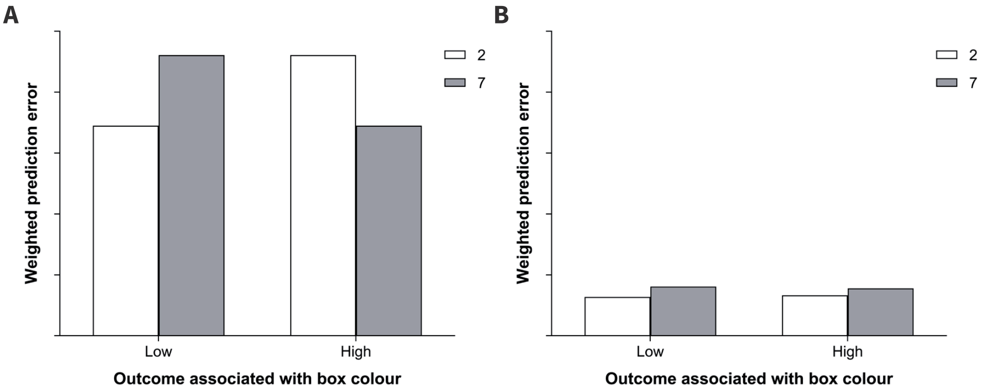


Figure A4. The expected reaction time (on a unit scale) for the four experimental conditions in Experiment 2, based on the weighted prediction errors at B_{12} (A) and at B_{23} (B).

Conclusion

We presented a computational characterisation of hierarchical predictive processing to deepen our understanding of the qualitative pattern of the results of the experiments. The pattern of results for the experiments shows a 3-way interaction for Experiment 1 and 3, with an interaction between agent and outcome only for the agent (or colour) question, and no such interaction for Experiment 2. This is in line with the patterns shown in Figure A3 and A4. Hence, the results can be explained with the proposed characterisation of hierarchical predictive processing.

3

Processing of prediction errors in mentalizing areas

Based on: Heil, L., Hartstra, E., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (submitted). Processing of prediction errors in mentalizing areas.

ABSTRACT

When seeing people perform actions, we are able to quickly predict the action's outcomes. These predictions are not solely based on the observed actions themselves, but utilize our prior knowledge of others. It has been suggested that observed outcomes that are not in line with these predictions result in prediction errors, which require additional processing in order to be 'explained away'. However, there is no consensus on whether this is indeed the case for the kind of high-level social-cognitive processes involved in action observation. In this fMRI study, we investigated whether observation of unexpected action outcomes causes additional activation in line with the processing of prediction errors and if so, whether this activation overlaps with activation in brain areas typically associated with social-cognitive processes. In the first part of the experiment, participants watched animated movies of two people playing a bowling game, one experienced and one novice player. In cases where the player's score was higher or lower than expected based on their skill level, there was increased BOLD activity in areas that were also activated during a theory of mind task that participants performed in the second part of the experiment. These findings are discussed in the light of different theoretical accounts of human social-cognitive processing.

INTRODUCTION

In our day-to-day social interactions, we collect information about people around us. For instance, we might remember that one person is very knowledgeable, while another person is good at playing tennis. We use this information to build up expectations of the behaviour of other people and the outcomes of their behaviour. We will therefore be more likely to ask our knowledgeable friend to join our pub quiz team, as this person will probably answer more questions correctly. This way, prior knowledge about others helps us understand the world and correctly respond to events.

This importance of prior knowledge is in line with the idea that predictions play a key role in cognitive processing. For instance, it has been suggested that predictions aid efficient processing as expected events are perceptually facilitated while unexpected events result in prediction errors, which require additional processing in order to be 'explained away' (Clark, 2013b; Friston, 2010; Rao & Ballard, 1999; Summerfield & De Lange, 2014). That is, when the unexpected events might be explained by a priori unlikely causes, then the prediction error can be minimized by assuming that such an unlikely cause indeed is in place, as this best explains the unlikely event. For rather low-level cognitive processes, such as visual perception, unexpected events have indeed been found to result in additional processing as indexed, for instance, by increased reaction times (Berti & Schröger, 2004; O'Reilly et al., 2013) and brain activity (Kok et al., 2012; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008).

Although there is no consensus that prediction errors also arise when people process more abstract, social information, several findings do suggest that this is the case. Koster-Hale and Saxe (2013) review evidence that both neural and behavioural responses to observed actions depend on whether this action is expected or unexpected. For instance, reaction times slow down when people observe another person holding an object with an incorrect grip (Bach, Knoblich, Gunter, Friederici, & Prinz, 2005; Van Elk, Van Schie, & Bekkering, 2009). In addition, areas that are part of the brain's mentalizing network show increased activity in response to unexpected actions. The mentalizing network includes the temporoparietal junction (TPJ), the superior temporal sulcus (STS) and the medial prefrontal cortex (mPFC) and is assumed be involved in mental state inference (Frith & Frith, 2006). Top-down and bottom-up signals in both TPJ and mPFC are modulated by the probability of an agent-caused event (Van Pelt et al., 2016) and brain activity in these and related areas increases when observed actions are not in line with a person's facial expression (Vander Wyk, Hudac, Carter, Sobel, & Pelphrey, 2009), with the properties of the object on which the action is performed (Bach, Gunter, Knoblich, Prinz, & Friederici, 2009; De Lange et al., 2008), with the request to cooperate (Shibata, Inui, & Ogawa, 2011), or when observed actions are implausible given the context (Brass et al., 2007).

In most of these studies, actions and their outcomes were expected or unexpected based on general world knowledge about the characteristics of objects and constraints imposed by certain contexts rather than information about specific persons. However, similar effects were found in studies focusing on situations in which people build up expectations based on their prior experiences with a specific person. For example, in studies in which participants read descriptions of another person's behaviour that were consistent (e.g. 'Tolvan gave her sister a hug') or inconsistent (e.g. 'Tolvan gave her mother a slap') with a trait suggested by a previous description (e.g. 'Tolvan gave her brother a compliment'), inconsistent descriptions resulted in increased brain activity compared to consistent descriptions (Dungan et al., 2016; Ma et al., 2011; Mende-Siedlecki et al., 2012). In these studies, however, participants did not observe unexpected actions online, but could imagine them based on stories describing the actions. It is unknown whether the same effects also arise in case people use prior knowledge about a specific person to predict the outcome of an observed action.

The studies described so far suggest that there might be a general mechanism for the processing of behaviour that is unexpected given prior knowledge a specific person, involving the explaining away of prediction errors. In case there is indeed such a general mechanism, then we might expect that the processing of observed action outcomes that are unexpected given our prior experiences with the person performing the action would also result in increased processing. Indeed, in the study discussed in **chapter 2**, we found that responses to questions slow down when people see that the outcome of another person's action is not as would be expected based on prior experiences, suggesting that prediction errors play a role in the processing of these events. As this was a behavioural study, we could not determine in which brain areas this additional processing takes place.

Therefore, we set out to test whether the processing of action outcomes that are expected or unexpected based on our knowledge about a person also causes additional activation in line with the processing of prediction errors. Furthermore, we investigated whether such additional activity would arise in brain areas encompassing the mentalizing network, which is associated with higher social-cognitive processes. In case brain activation related to unexpected action outcomes would be found to overlap with activation in brain areas typically associated with social-cognitive processes, this would support the idea that prediction errors do not only arise at lower, more perceptual levels, but also at higher, more abstract levels.

In this fMRI study, participants watched animated movies of two people playing a bowling game, one experienced and one novice player. Assuming that unexpected events would result in prediction errors that require additional processing in order to be explained away, we hypothesized that in cases where the player's score was higher or lower than expected based on their skill level BOLD activity would increase. Moreover, based on the idea that prediction errors do not only drive low-level, but also more high-level social-cognitive processing, it was expected that this additional activity would also arise in brain areas traditionally associated with social-cognitive processes.

METHODS

Participants

Thirty-five healthy, right-handed individuals with normal or corrected-to-normal vision participated in this study. Data from twelve participants were excluded: in one case because of low accuracy (< 85%) and in the other cases because participants were unable to comply with the task demands, for instance due to early discontinuation of the experiment or excessive head motion. This resulted in a final dataset of 23 participants (14 women, 9 men) aged between 18 and 28 (M = 21.91 years, SD = 2.47 years). All but one participant scored below the cut-off of 32 on the Dutch translation of the Autism Spectrum Quotient questionnaire (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Hoekstra, Bartels, Cath, & Boomsma, 2008). Written informed consent was obtained from each participant and participants received course credits or twenty euro for participation. The study was approved by the local ethics committee (CMO Regio Arnhem-Nijmegen).

Stimuli

For the bowling task, 24 animated movies were created using Autodesk's 3ds Max 2014 and MotionBuilder 2014. Each movie showed one out of two possible bowling players (selected from WorldViz Vizard Complete Characters) on a bowling lane, throwing a ball directed at the pins at the end of the lane. The ball rolled either slightly left or slightly right of the centre and upon hitting the pins, 1, 2, 3, 6, 7 or 8 pins fell down. The kinematics of the ball movement were not related to a specific player or outcome. Each movie lasted 5 seconds and the player disappeared after 1.2 seconds, in order to keep the visual display of the action outcomes the same for the two players.

For the theory of mind task, we used the localizer task described by Dodell-Feder, Koster-Hale, Bedny, and Saxe (2011; http://saxelab.mit.edu/superloc.php), based on the initial task by Saxe and Kanwisher (2003). This task used twenty short stories either about a person holding a false belief or about an outdated representation, such as a map or photograph showing something that no longer exists. These stories were translated to Dutch, as all participants were native or fluent speakers of Dutch.

All stimuli were presented using Presentation software (version 17.2, www.neurobs.com).

Bowling task

For the bowling task, we used a within-subject design that was similar to the one in the study described in **chapter 2**. As in that study, participants read instructions on the screen, explaining that they would be watching movies of two people playing a bowling game and answer questions about these movies. They were told that there was one novice and one experienced player, who usually obtained scores matching their skill levels. In order to allow them to associate appearance of the agent with his skill level, participants performed four practice trials, in which they received information about which agent they would see before each movie was presented.

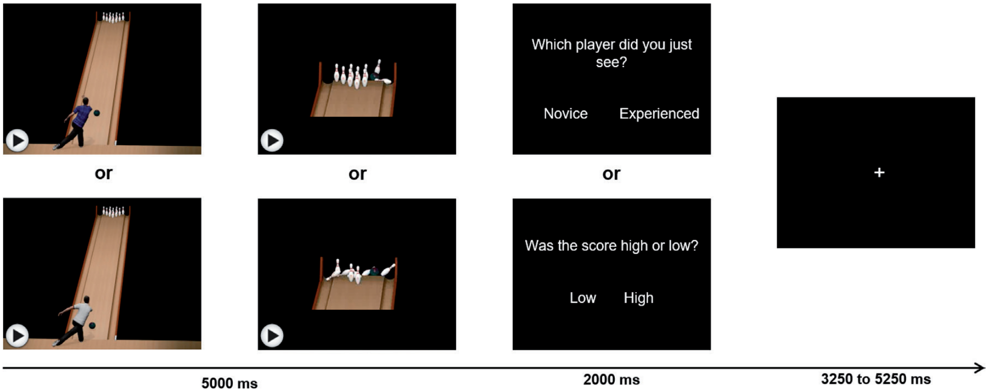


Figure 1. Schematic representation of the procedure in the bowling task. Each trail started with a 5000 ms bowling movie showing either the novice or experienced player who scored either high or low. This movie was followed by one out of two possible questions concerning the player or the score. The question was presented for 2000 ms, even if participants required less time to answer it. Each trial ended with a fixation cross.

In the main task there were 288 trials in total, all following the procedure shown in Figure 1. In 75% of all trials, movies showed the novice or experienced player obtaining an expected score, whereas in 25% of all trials they showed the player obtaining an unexpected score of at least 4 points higher or lower than their average score. More specifically, the novice player received a low score (1, 2, or 3) in 108 out of 144 trials, whereas the experienced player received a high score (6, 7, or 8) in 108 out of 144 trials. Within the category of low scores, a score of 2 was most frequent (96 out of 144 trials), as was a score of 7 in the category of high scores. The other scores (i.e. 1, 3, 6, and 8) appeared very infrequently (a total of 32 out of 288 trials) and were included as fillers that provided variability in scores, in order to make the experiment more naturalistic.

After each movie, participants answered one out of two questions. These questions were written in Dutch and either asked whether they saw the experienced or the novice player or whether a small or a large number of pins fell down. These questions were presented in random order and participants were not aware which question they would be asked in each trial. Each question was presented on the screen for two seconds. Participants were instructed to answer as quickly as possible by pressing a left or a right button on a button box. These buttons corresponded to two answer options presented underneath the question, the order of which was randomized in order to prevent motor preparation. Unlike in the practice trials, participants did not receive feedback on the accuracy of their answer. Each trial was followed by a fixation cross, presented for a duration randomized between 3250 and 5250 milliseconds. Data were collected in one run and participants could take a short break in the scanner after the first 144 trials.

Theory of mind task

We used an adapted version of the theory of mind (ToM) task that is commonly used to identify brain areas specific for mentalizing (Dodell-Feder et al., 2011; Saxe & Kanwisher, 2003). In this task, participants read a total of twenty short stories, describing either a false belief or a false photograph. Following the procedures used by Dodell-Feder and colleagues, each story was presented for 10 seconds and followed by a question about the story. Participants were given 4 seconds to answer the question before the trial ended. The order of the stories was randomized for each participant. After each question, a fixation cross was shown for 12 seconds (see Figure 2).

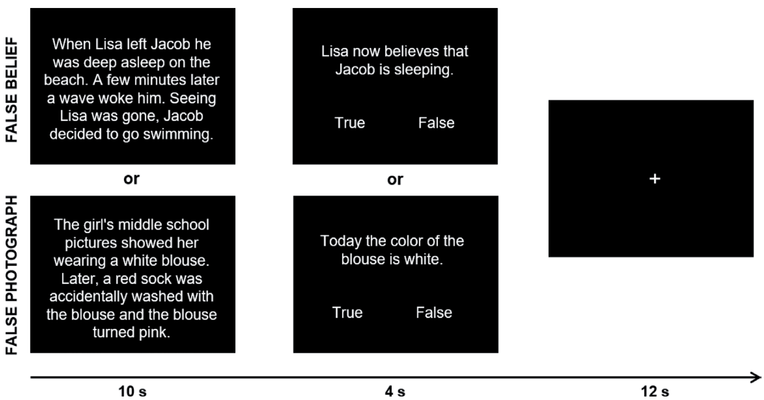


Figure 2. Schematic representation of the procedure in the ToM task, with examples from the original version of the task in English. Each trial started with a short story that was followed by a statement. Participants indicated whether the statement was true or false. Each trial ended with a fixation cross.

fMRI data acquisition

Functional images were acquired using a 3-Tesla MRI scanner (Siemens Magnetom Skyra, Erlangen, Germany) with a 32-channel head coil. A multi-echo EPI sequence sensitive to the blood-oxygen-level-dependent (BOLD) signal contrast was used (34 transversal slices with a voxel resolution of 3.5 × 3.5 × 3.0 mm, acquired in ascending order, interleaved scanning, TR = 2.07 s, TE = 9 ms, 19.25 ms, 29.5 ms, and 39.75 ms, distance factor = 17%, 90° flip angle, 224 mm FOV). The first 30 volumes were used for calculation of the echo-weighting parameters and were discarded in data analysis. During the break between the bowling task and the ToM task, high-resolution anatomical images (voxel size 1 × 1 × 1 mm) were acquired using a T1-weighted MP-RAGE sequence (TR = 2.3 s, TE = 3.03 ms, flip angle 8°, 256 mm FOV).

Preprocessing

Data for both the bowling task and the theory of mind task were preprocessed and analysed using SPM8 software (http://www.fil.ion.ucl.ac.uk/spm; Wellcome Trust Centre for Neuroimaging, London, UK) running on Matlab 2012b. First, before combining the four echoes, all the first echo volumes were realigned to the first volume of the first echo and the volumes of the remaining echoes were realigned to the first echo and resliced. To combine the echoes within each volume, the echo-weighting parameters derived from the first 30 volumes were applied to the echoes of all remaining volumes. In this process, a mean functional image and head movement parameters were created as well. Next, functional images were temporally aligned to the middle slice of each volume. The mean functional image was co-registered to the anatomical image and the anatomical image was segmented into white matter, grey matter and cerebrospinal fluid. The resulting normalization parameters were used to normalize the anatomical and functional images to the standard template in Montreal Neurological Institute (MNI) space. Finally, the functional images were spatially smoothed using a Gaussian kernel of full-width 8 mm at half maximum.

fMRI analyses

Preprocessed fMRI data were analysed using a General Linear Model (GLM) as implemented in SPM8. For each participant, we analysed BOLD data as a function of condition. For the bowling task, we distinguished four conditions: 2 levels of expectancy (expected vs. unexpected outcome) × 2 questions (agent vs. outcome question). In all four conditions, the onset was set to the end of the bowling animation, when the question appeared on the screen. We included six rigid-body motion parameters obtained during preprocessing, errors and onsets of button presses in the model as regressors of no interest. Data were convolved with a canonical hemodynamic response function (HRF) as well as the first time derivative, of which only the canonical hemodynamic response was used for further analysis. Contrasts for each of the four conditions, derived from these participant-specific first-level analyses, were introduced to a second-level random-effects analysis, in order to allow for population inferences. This was a full-factorial whole-brain analysis, in which the main effect of expectancy was analysed by subtracting brain activity in the expected trials from that in the unexpected trials. In this analysis, the trials related to the two questions were combined and not analysed separately.

For the ToM task, we distinguished two conditions: false belief and false photograph. The onset was set to the moment the question appeared on the screen. In the analysis, we again included six rigid-body motion parameters as regressors of no interest. Data were convolved with a canonical HRF and first-level contrasts (false belief > false photograph) were derived. These contrasts were introduced to a second-level analysis, in which a t-test was used in order to analyse the additional brain activity related to processing false beliefs as compared to false photographs. In a follow-up analysis, we created a mask of the false belief > false photograph contrast from the ToM task. Using inclusive masking, we then analysed whether areas within this mask showed up in the unexpected > expected contrast of the bowling task.

For all analyses, we used a cluster-level familywise error (FWE) rate of 0.05, with clusters of activation defined with a voxel-level p-value of 0.001 (uncorrected).

Behavioural analyses

For the bowling task, reaction times to the questions that followed movies in which the outcome was 2 or 7 were analysed using a 2 (expected vs. unexpected) × 2 (agent question vs. outcome question) repeated measures ANOVA. Reaction times to questions following movies with scores 1, 3, 6, and 8 and those to questions that were answered incorrectly were excluded from the analysis. The average accuracy was 97.2% over all trials (96.8% for the agent question and 97.6% for the outcome question). On average, only 1.3% of trials was not answered within the time limit and thus counted as missing.

For the ToM task, accuracy data suggest that the task was sufficiently difficult. Participants scored on average 70.7% correct over all trials. This number was rather similar for the false belief (70.9%) and the false photograph (70.4%) statements. Behavioural data for this task were not further analysed, as there were no relevant questions related to these data.

RESULTS

Behavioural data bowling task

Based on the assumption that unexpected events result in prediction errors that require additional processing, reaction times to questions following unexpected events (i.e. a novice player obtaining a high score or an experienced player obtaining a low score) were anticipated to be higher than those to questions following expected events (i.e. a novice player obtaining a low score or an experienced player obtaining a high score). The results of the bowling task show that this is indeed the case: participants needed more time to answer questions that followed an unexpected outcome (M = 924.20, SD = 168.16), compared to questions that followed an expected outcome (M = 899.62, SD = 165.08), $F(1, 22) = 5.23$, $p = .03$, $\eta_p^2 = .19$. In addition, there was a significant effect of question,

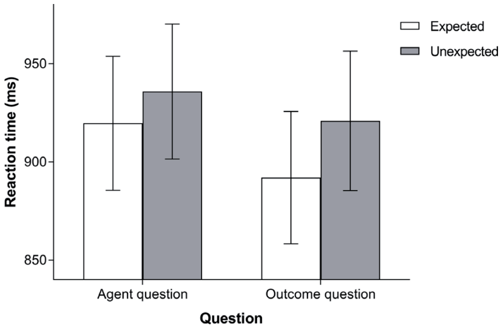


Figure 3. Reaction times (mean ± SEM) to questions about the agent or the outcome following scores that were expected (white bars) or unexpected (grey bars) given the agent's skill level.

$F(1, 22) = 25.39$, $p < .001$, $\eta_p^2 = .54$, with longer reaction times for the agent question (M = 930.09, SD = 149.74) than for the outcome question (M = 861.59, SD = 159.78) but no significant interaction effect between question and expectancy, $F(1, 22) = 3.56$, $p = .07$, $\eta_p^2 = .14$ (see Figure 3).

fMRI data bowling and ToM task

In the fMRI analyses of the bowling task, we tested whether the observation of unexpected action outcomes causes additional activation as compared to expected outcomes, in line with the processing of prediction errors. In this whole-brain analysis, data for the question about the outcome and the question about the player were combined as the questions mostly served to ensure that participants would attend to all aspects of the scene and no interaction effect between question and expectancy was found in the behavioural analysis. Our main interest was in the unexpected > expected t-contrast, which revealed significant activation in the left and right inferior parietal lobule, left and right middle frontal gyrus, left middle temporal gyrus, and right superior frontal gyrus (see Table 1 and Figure 4).

Table 1. Activations for the unexpected > expected contrast. All results are cluster-level corrected using a FWE-rate of 0.05. MNI coordinates (x, y, z) of local maxima for each cluster are given, including the cluster voxel extent (k) and the t-statistic (t) at those coordinates.

Brain region	x	y	z	k	t
Right Inferior Parietal Lobule	51	-61	46	332	6.13
Left Middle Frontal Gyrus	-42	50	-11	321	5.54
Left Inferior Parietal Lobule	-45	-64	46	520	5.50
Left Middle Temporal Gyrus	-57	-34	-14	188	5.22
Left Middle Frontal Gyrus	-42	14	43	189	5.02
Right Middle Frontal Gyrus	48	29	40	130	4.29
Right Superior Frontal Gyrus	12	44	52	117	3.91

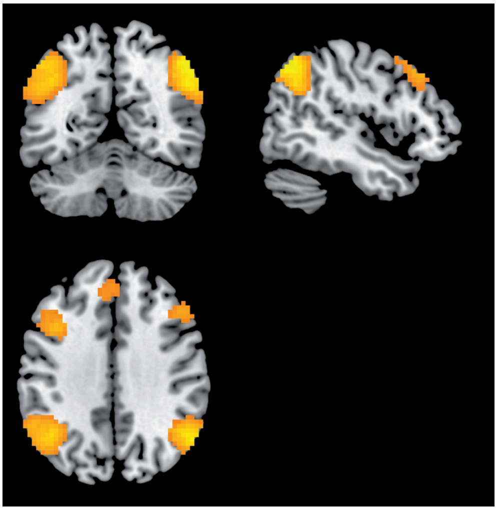


Figure 4. BOLD responses following unexpected events as compared to expected events (unexpected > expected) in the bowling task. For visualization purposes, only significant clusters are shown.

In the analyses of the ToM task, activation arising when participants read a story describing a false photograph was subtracted from the activation arising when they read a story describing a person holding a false belief. This false belief > false photograph contrast showed significant activation in the left superior temporal gyrus, left and right middle temporal gyrus, cingulate gyrus, right inferior parietal lobule, left and right medial frontal gyrus, and right inferior frontal gyrus (see Figure 5 and Table 2).

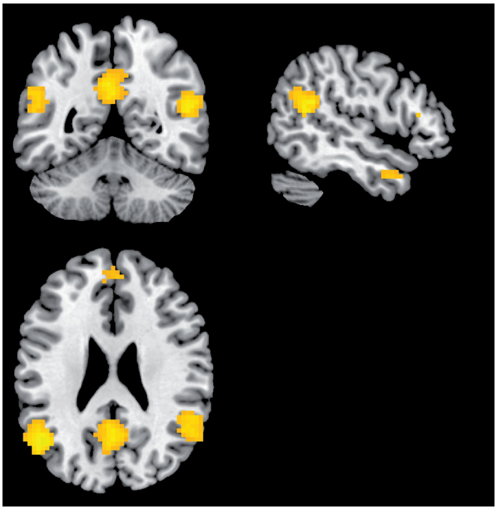


Figure 5. BOLD responses following false belief as compared to false photograph stories (false belief > false photograph) in the ToM task.

Table 2. Activations for the false belief > false photograph contrast. All results are cluster-level corrected using a FWE-rate of 0.05. MNI coordinates (x, y, z) of local maxima for each cluster are given, including the cluster voxel extent (k) and the t-statistic (t) at those coordinates.

Brain region	x	y	z	k	t
Left Superior Temporal Gyrus	-48	-61	25	202	8.83
Left Middle Temporal Gyrus	-51	-22	-8	33	8.19
Cingulate Gyrus	0	-58	28	276	8.17
Right Inferior Parietal Lobule	51	-49	22	162	8.10
Left Middle Temporal Gyrus	-45	8	-29	16	7.73
Left Medial Frontal Gyrus	-3	50	31	45	7.68
Right Middle Temporal Gyrus	48	5	-29	23	7.24
Left Medial Frontal Gyrus	-3	56	-5	22	7.23
Right Inferior Frontal Gyrus	51	26	13	2	7.09
Right Medial Frontal Gyrus	15	47	4	1	6.53
Right Middle Temporal Gyrus	60	-7	-17	3	6.31

Furthermore, we analysed whether additional activation related to unexpected events in the bowling task partly overlaps with activation in brain areas typically associated with social-cognitive processes, using the inclusive mask based on the functional data from the ToM task. This mask, based on the false belief > false photograph contrast, was assumed to reflect activity associated with social-cognitive processing. Using inclusive masking, we analysed whether there were areas within this mask that were significantly more active during the observation of unexpected as compared to expected events in the bowling task. As in the previous analysis, we analysed the main effect of expectancy by subtracting brain activity in the expected trials from that in the unexpected trials. The results (presented in Table 3 and Figure 6) show activation in both the left and the right angular gyrus, including the left and right TPJ.

Table 3. Activations for the unexpected > expected contrast, masked with the activation from the false belief > false photograph contrast from the ToM task. All results are cluster-level corrected using a FWE-rate of 0.05. MNI coordinates (x, y, z) of local maxima for each cluster are given, including the cluster voxel extent (k) and the t-statistic (t) at those coordinates.

Brain region	x	y	z	k	t
Right Angular Gyrus *	51	-58	34	120	5.56
Left Angular Gyrus **	-42	-58	34	213	4.65

* This cluster includes the right temporoparietal junction (rTPJ)

** This cluster includes the left temporoparietal junction (lTPJ)

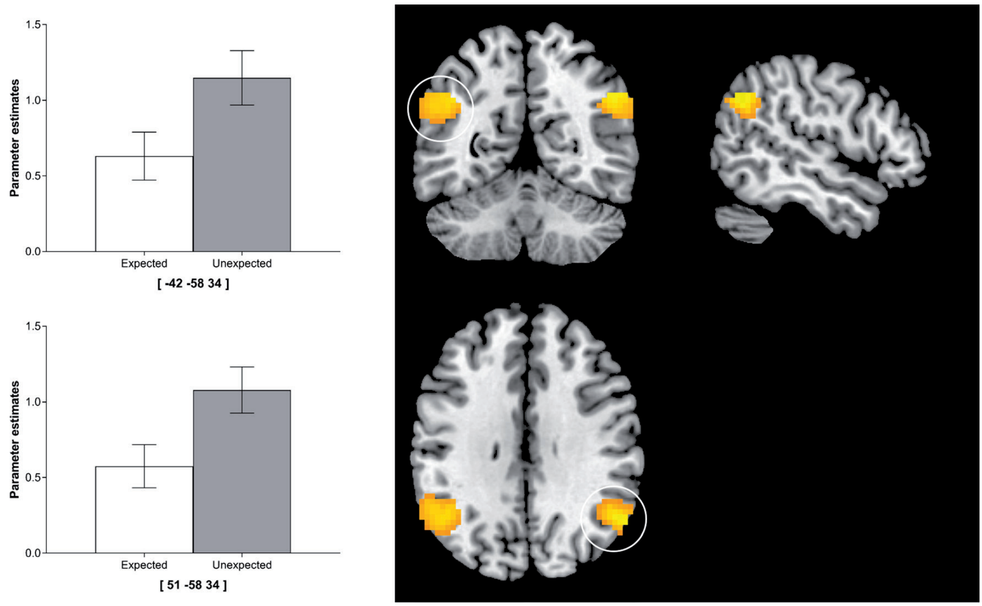


Figure 6. Overlap in BOLD responses between the unexpected > expected contrast in the bowling task and the false belief > false photograph contrast in the ToM task. Only significant clusters are shown. Parameter estimates of activity in the left and right angular gyrus are shown in the panels on the left.

DISCUSSION

In this study, we investigated the processing of action outcomes that are expected or unexpected given prior knowledge about the person performing the action. There are two main findings. First, the results of this experiment show increased neural activity in the inferior parietal lobule, middle and superior frontal gyrus, and the middle temporal gyrus when participants saw a bowling player obtaining a higher or lower score than expected. Second, comparing the activity arising during this bowling task with that arising during a theory of mind task, we found that in an area around both the left and right angular gyrus, which includes the left and right TPJ, neural activity increased not only when participants observed unexpected action outcomes, but also when they considered the false beliefs of another person.

The finding that the TPJ shows increased activity during the processing of unexpected action outcomes as well as during the processing of other people's false beliefs suggests that observation of action outcomes that are expected or unexpected given previous experiences with the person performing the action involves higher level social-cognitive processes. The TPJ has been pinpointed as one of the core areas in the mentalizing network, a network of areas involved in the processing of social information. As this area is not only recruited during the bowling task used in this experiment, but also when people read stories about the thoughts or behaviour of other people, it seems that online observation of agent's actions and their consequences activates knowledge or processes also engaged when mentalizing about imagined scenarios.

Some have suggested that the TPJ is specifically involved in thinking about another person's thoughts (Saxe & Powell, 2006), although the area has also been found to be active during more general attention-related tasks, which has inspired alternative theories about its functions. For instance, it has been suggested that there are different clusters within the TPJ, one involved in attention, the other in social processing (Krall et al., 2015; Mars et al., 2011). It has also been suggested that the area may be primarily involved in reorienting attention and that it is also active during tasks requiring social cognition, because such tasks involve reorienting of attention and other more domain-general processes (Corbetta, Patel, & Shulman, 2008; Decety & Lamm, 2007; Mitchell, 2007; Van Overwalle & Baetens, 2009). Our findings do not distinguish between these different theories and do therefore not imply that the processes involved are necessarily uniquely social.

Importantly, however, the action outcomes are only unexpected based on participants' prior experiences with the bowling player: knowing that a player is experienced, makes a low score unexpected. The outcomes themselves are equally frequent and our results can therefore not be explained by an overall difference in frequency. As we have suggested in the study described in **chapter 2**, it seems that the bowling task in this study not only requires associating a colour and a score, but also considering the causal relation between the two. Knowing that the player causes the ball to move towards the pins and knock them over allows people to predict the score based on which person is currently playing. This suggests that the areas activated are involved in more abstract, higher-order processes that support predictions about action outcomes based on knowledge about a person.

The finding that observation of action outcomes that are unexpected given previous experiences with the person performing the action causes increased neural activity is in line with the idea that prediction errors play a key role in cognitive processing. As such, it is also in agreement with previous studies showing that the processing of unexpected events requires more resources in terms of time (e.g. O'Reilly et al., 2013) and neural activation (e.g. Summerfield et al., 2008). A potential interpretation of these findings is that there is a common mechanism involving the explaining away of prediction errors that underlies not only the processing of observed behaviour that is unexpected given prior knowledge about that specific person, but the processing of unexpected events in general. The idea of such a general mechanism is a central assumption of the predictive processing framework, which has been posed as a unifying framework for brain functioning (Clark, 2013b; Friston, 2005). In this framework, it is assumed that incoming information is compared to predictions and that potential prediction errors arising from this comparison are then used to update the generative models underlying these predictions. In this sense, cognitive processing is primarily focused on this explaining away of prediction errors. Although the idea that this framework can explain perceptual inference is gaining acceptance, there is no consensus on whether it can also explain more abstract, social processing (e.g. Koster-Hale & Saxe, 2013). Our findings are in line with this last idea, but they do not distinguish between the predictive processing and other frameworks. For instance, based on an account in which probabilistic inference takes place after rather than before the observation of the events, one could assume that inference of a less probable event requires more processing resources than inference of a more probable event, resulting in increased activity in trials that we labelled as 'unexpected'. Based on such an account we would indeed predict similar results, but whereas the predictive processing framework assumes that this additional processing is required in order to explain away prediction error, it is unclear which mechanism requires additional processing in case of such non-predictive type of probabilistic inference. Since the posterior probability of the causes is conditioned on the observed event it is computationally no longer relevant whether the event was likely or unlikely.

Furthermore, it seems that the predictive processing framework could actually be seen as complementary to several other accounts of human cognition. Especially associative theories are sometimes suggested to compete with the predictive processing framework, but they might not necessarily oppose it. As Press, Heyes, and Kilner (2011) suggest, for example, the associative sequence learning model and the predictive processing framework seem to address related but different questions. Associative sequence learning, on the one hand, explains how relations are learned whereas predictive processing, on the other hand, explains how learned relations support inferences about other people's actions and their outcomes. Moreover, associative sequence learning assumes that events are associated when their occurrence follows the principles of the Rescorla-Wagner model (Cooper et al., 2013). According to this model, learning does not take place because two events simply co-occur, but because this co-occurrence is unexpected and thus results in a prediction error (Rescorla & Wagner, 1972).

In sum, the findings of the current study are in line with the idea that prediction errors play a key role in the processing of action outcomes that are unexpected given prior knowledge about a person and that part of these prediction errors is explained away in brain areas typically associated with social-cognitive processes. Future research will need to determine whether social-cognitive processes are indeed built upon general predictive principles that have been previously attributed to more low-level processing.

4

Biased person performance models: how ingroup overestimation persists in the face of outgroup individuation

Based on: Heil, L., Brecht, S., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (submitted). Biased person performance models: How ingroup overestimation persists in the face of outgroup individuation.

ABSTRACT

Our judgments of another person's performance or characteristics may be biased depending on whether the person is considered part of one's own social group (the ingroup) or another social group (the outgroup). For instance, people sometimes overestimate performance of ingroup members compared to outgroup members and distinguish less between individual outgroup than ingroup members. Little is known, however, about the models underlying predictions about individual task performance. Do these person performance models differ between ingroup and outgroup? And do they distinguish between individuals in both ingroup and outgroup? We investigated this using a minimal group paradigm in which participants predicted the scores of players from two different teams in a bowling game. Participants' predictions showed clear individuation in both outgroup and ingroup (i.e. they distinguished between individual members in both groups), but also a reliable overestimation of ingroup performance. This suggests that ingroup overestimation occurs even in conditions where reliable information about task performance is available and both ingroup and outgroup members are clearly individuated. In other words, ingroup overestimation seems to be a robust bias that does not disappear when information about a person's actual performance is provided. This would imply that ingroup overestimation is not simply the result of having more detailed knowledge about one's own social group and that gaining more information about outgroup members does not necessarily reduce ingroup overestimation. We discuss the societal implications of these findings.

INTRODUCTION

Whether judging if someone is a good fit for a certain job, if one sports player is better than the other, or if someone is a friendly person, we often make judgments about another person's performance or characteristics. Rather than being purely objective, these judgments are biased depending on the social group the person belongs to. One prominent example of this is that our judgments are, consciously or unconsciously, influenced by whether the person is considered part of one's own social group (the ingroup) or another social group (the outgroup).

For instance, when encountering people that we consider to be members of an ingroup, we tend to like them more than others (e.g. Martiny-Huenger, Gollwitzer, & Oettingen, 2014; Nesdale, Maass, Griffiths, & Durkin, 2003) and associate them with more positive character traits (Doise et al., 1972; Otten & Moskowitz, 2000). This so-called 'ingroup bias' occurs automatically (Ashburn-Nardo, Voils, & Monteith, 2001) and is associated with 'ingroup favouritism': a tendency to act more prosocially towards members of our own social group (Everett, Faber, & Crockett, 2015; Tajfel, 1970; Turner, 1975).

When judging other people's actions, this ingroup bias seems to express itself as an overestimation of ingroup performance. One of the few studies reporting on this is a classic study by Hastorf and Cantril (1954). In their experiment, students watched a football match between two teams and they reported fewer infractions for a certain team when this team was their own rather than the rivalling team. In a more recent study by Molenberghs and colleagues (2013), participants performed a task in which they moved their hands as quickly as possible. Afterwards, when they viewed movies of ingroup and outgroup members performing the same task, they judged the movements of ingroup members to be faster than those of outgroup members.

In addition to this overestimation of ingroup characteristics and performance, it seems that people distinguish more between ingroup members than between outgroup members. The most prominent example of such a difference in individuation is that people have been found to be better at recognizing faces of members of their own group, even when they are randomly assigned to this group and perceptual expertise is the same for both groups (Bernstein et al., 2007; Van Bavel, Swencionis, O'Connor, & Cunningham, 2012). Possibly, this is the result of increased attention for information related to ourselves and our social group, which would be in line with the finding that people are faster in judging information related to themselves (Sui, He, & Humphreys, 2012) and with neuroimaging studies showing increased neural responses to members of an ingroup and an outgroup. For instance, viewing faces of members of an ingroup was found to be associated with a larger amplitude of the face-sensitive N170 ERP component (Golby et al., 2001) and more activation in the fusiform gyrus and other brain areas (Ratner & Amodio, 2013; Van Bavel et al., 2008). Furthermore, Molenberghs et al. (2013)

found increased activity in the left inferior parietal lobule when people observed hand actions of their own team as compared to those of the other team.

It is conceivable that increased individuation and overestimation of ingroup members as compared to outgroup members also influences the way people form expectations about the future performance of others. In order to form such expectations, people need to build cognitive models based on prior experiences. Some researchers suggest that such generative models not only allow us to make predictions, but that they also guide perception (Clark, 2013b; Friston, 2005). According to this predictive processing view, not only basic visual perception, but also social perception is best described as a process of top-down hypothesis testing (e.g. Bach & Schenke, 2017; Koster-Hale & Saxe, 2013; Otten et al., 2017). Although person performance models are indeed likely to be used as generative models that drive perception, in this study we are interested in the way in which these models allow us to predict the performance of people around us.

Little is known, for instance, about the way in which the models that we use to predict the performance of ingroup members differ from those we use to predict the performance of outgroup members. Given the evidence on ingroup overestimation and individuation, there are two possibilities, which are not mutually exclusive. First, one might expect the person performance models for groups to differ in the sense that members of our own group are represented more positively than members of the other group, resulting in a general overestimation of their characteristics and performance. In other words, the models may reflect ingroup overestimation. Second, the person performance models may differ in the sense that the model for the ingroup is based on information about individual task performance, whereas the model for the outgroup is based on information about task performance on a group level. In this case, the person performance models would reflect a difference in individuation between ingroup and outgroup.

It is not unlikely that people's generative models are both more positive and more individuated for ingroup members. As far as we are aware, however, previous studies have neither specifically investigated the cognitive models that drive predictions about other people's performance, nor have they investigated overestimation and individuation at the same time in one experiment. For instance, many studies showing increased individuation of ingroup members focus on face recognition rather than judgment of other people's task performance. As such, they do not test any kind of overestimation of task performance. In contrast, studies showing ingroup overestimation do not provide participants with reliable information about individual group members' task performance. In the study by Molenberghs et al. (2013), for example, participants only saw a hand and could not identify individual group members. Hence, such studies do not test individuation.

In this study, we aimed to investigate the models underlying predictions when reliable information about individual task performance is available. We used a minimal group design in which participants were randomly assigned to one of two teams. They viewed animated movies of people playing a bowling game and were instructed to predict the score in each movie. These predictions could be based on the average score of a specific player. In total, there were ten players, divided over two teams. We hypothesized that the predictions of participants would indicate that they distinguish more between members of the ingroup and generally overestimate their performance.

In the following section, we disclose all measures, manipulations, and exclusions, as well as the method for determining the final sample size.

METHODS

Participants

Thirty healthy participants (mean age: 21.30 ± 2.27, all women) with normal to corrected-to-normal vision participated in the study. The sample size of 30 participants was determined before data analysis and we continued to test until we reached this goal. No participants were excluded from the analyses. We only invited women to participate, as we hoped this would aid identification with the female players in the stimulus movies. The participants received course credits or 10 euro for participation. The study was approved by the institution's local ethics committee and written informed consent was obtained from each participant.

Stimuli

Using Autodesk's 3ds Max 2014 and MotionBuilder 2014 (www.autodesk.com), animated movies with a duration of 4500 ms were created. These movies showed a bowling player throwing a ball directed at ten pins standing at the end of the bowling lane. Two different avatars were used for the bowling players: one with a blue shirt and one with a red shirt. Both were taken from Worldviz Vizard Complete Characters (www.worldviz.com/products/avatars/complete-characters) and were selected based on their female appearance. Half of the movies showed the ball rolling slightly towards the left side of the lane and the other half showed the ball rolling slightly towards the right side of the lane. In all movies, the ball knocked over at least one pin. This resulted in 36 movies, differing in terms of the colour of the player's shirt (blue or red), score (1 to 9) and trajectory of the ball (left or right).

The movies were used to represent the actions of ten bowling players. These players could be identified by pictures selected from the Radboud Faces Database (Langner et al., 2010). All pictures showed a young female face with a neutral facial expression. Two independent ratings of attractiveness were used to ensure similarity of the faces.

Procedure and design

When participants entered the lab, the experimenter informed them that they were assigned to either the red team or the blue team. They were instructed to wear a shirt and a wristband in the team's colour. In addition, a logo in the same colour was placed next to the screen. The coloured cues were used to create a more realistic setting that enabled identification with the own team and to remind participants of their team affiliation during the entire task.

After the group manipulation, participants were told that they would be watching movies that were based on actual performance of players of their team and the other team in a bowling competition and that they would be asked to predict their scores. They received no further explanation of why they needed to report these predictions. Furthermore, participants were instructed to observe each action carefully, as they would be asked which team won at the end of the task. In order to make the task more realistic and to facilitate group affiliation, they were also told that they would receive a reward in case their own team would win.

The task was presented using Presentation software (version 17.1, www.neurobs.com). It started with the calibration of an eye tracker, which was used to track participant's eye movements. We intended to measure the amount of time spent looking at the player, in order to analyse whether this would correlate with the amount of individuation within each team. For technical reasons, however, we were only able to collect eye tracking data for 22 out of 30 participants and we therefore decided not to include these data in our analyses.

After calibration, participants performed four practice trials, which were independent of the main task and served to familiarize participants with the task. During these practice trials, participants received feedback on their performance and case of insufficient performance on these trials, instructions were repeated.

In the main experiment, both teams consisted of five players, each with a different average score (or skill level) ranging from three to seven (Figure 1A). Before each movie, a picture of the current player was shown (Figure 1B). This picture was presented in a blue or red frame, indicating to which team the player belonged. Participants were asked to indicate which score they expected this player to get by pressing the corresponding number on a keyboard. The screen showing the picture and the question was shown for 2500 ms or until the question was answered.

After that, the movie was presented, showing the player receiving her average score or a score close to that. In total, there were 340 trials, presented in random order and equally distributed over the ten players. Of the 34 trials in total for each player, she received her average score in fourteen trials, a score of one point above or below average in sixteen trials, and a score of two points above or below average in four trials. In 23 of all 340 trials, the movie was followed by a question about which score participants just saw, to ensure that participants would attend to the actual outcome. After each trial, a fixation cross was presented for 700 ms.

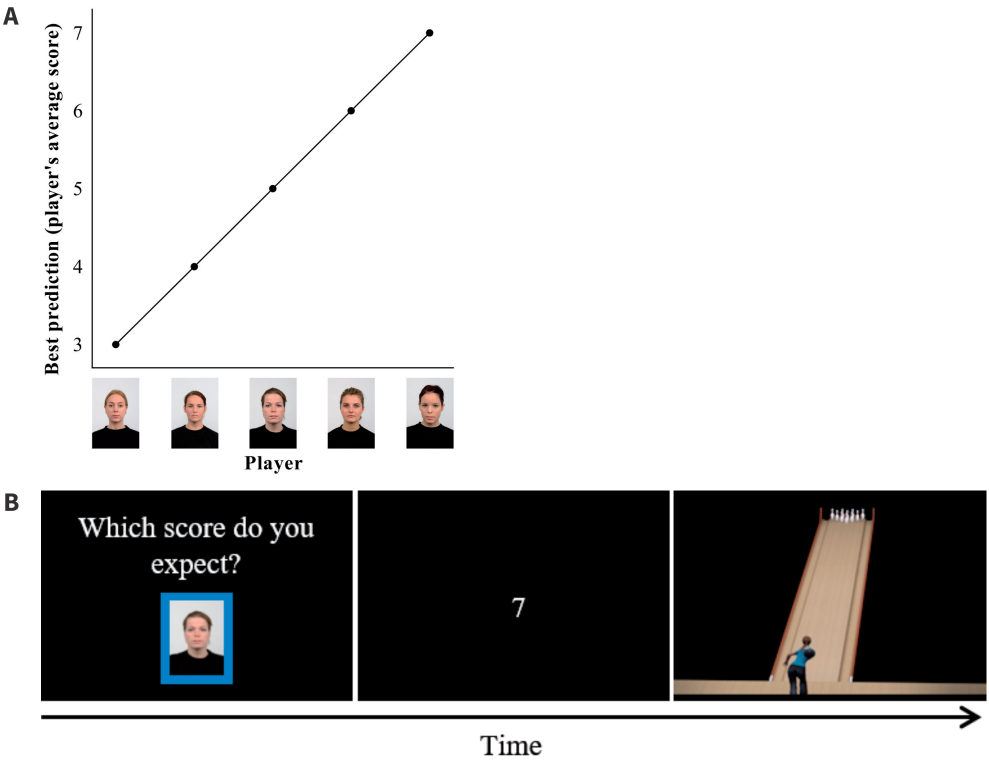


Figure 1. Example of players and their average scores for one of the teams (A) and a schematic overview the main events in the bowling experiment (B).

The bowling task was followed by four questions that served to check if participants indeed affiliated more with their own than with the other team and if they assumed that their team was better than the other. After these questions, participants performed an implicit test of team affiliation. This task was an adjusted version of the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) and was intended to test participants' associations between the teams and their members on the one hand and words related to self and other on the other hand. There were 260 trials, divided over five blocks of which three were merely practice blocks. In each block, participants categorized pictures of bowling players from the bowling task they performed previously and/or words related to 'self' and 'other', based on the words used in a previous study (Greenwald & Farnham, 2000). They were instructed to perform this categorization by pressing buttons on the left and right as quickly as possible, as the pictures and words disappeared after 2000 ms. The categories were 'blue team' versus 'red team' and 'self' versus 'other'. In the two main blocks, one button was used for two different categories. For example, in the congruent block, the same button was used to categorize pictures of members from the own team and self-related words. In the incongruent block, the same button was used to categorize pictures of members of the own team and other-related words. The slowing in reaction time in the incongruent block was then assumed to reflect the interference caused by the associations between the own team and self-related words and between the other team and other-related words. In other words, we assumed that the more participants affiliated with their team, the more they slowed down when they had to link their team to the concept of 'other'. As the IAT has been found to be insensitive to voluntary control (Kim, 2003), the results should reflect actual team affiliation, which is not influenced by, for example, demand characteristics.

At the end of the experiment, participants filled out a questionnaire to assess their competitive attitude (adapted from Ryckman, Hammer, Kaczor, & Gold, 1990). Finally, participants were told that their own team had won the bowling competition and they were all given a candy bar as a reward before leaving the lab.

RESULTS

Team affiliation

Before being able to interpret the results from the bowling task, it is important to analyse if participants indeed affiliated with their own team. They answered questions about how strongly they felt part of the red team and the blue team on a twenty-point visual analogue scale ranging from -10 to 10. As expected, a paired-samples t-test comparing the answers to these questions shows that participants felt part of their own team (M = 4.67, SD = 2.51), more than the other team (M = 2.80, SD = 1.63), $t(29) = 3.39$, $p = .002$.

This is in agreement with findings from the IAT, which provides a more implicit measure of team affiliation. For the analysis of this task, trials with very short reaction times (< 100 ms) and trials in which no answer was given within the time limit of 2000 ms were excluded. Reaction times for incongruent blocks (M = 803.81, SD = 120.72), were significantly higher than for congruent blocks (M = 726.83, SD = 104.08), $t(29) = 4.55$, $p < .001$. This finding suggests that participants associated themselves more strongly with members of their own team than with members of the other team and that this association slowed down their responses when the task required them to follow instructions that were incongruent with this association.

In the first block of the IAT, participants matched pictures of the players to the teams. Although they only had 2000 ms for each picture, participants managed to do this correctly in on average 96.7% (and minimally 85%) of all trials. This suggests that they learned to associate the players with the correct teams.

Bowling task

The predicted scores in the bowling task were analysed using a 2 (teams: own vs. other) × 5 (skill level: 3, 4, 5, 6, or 7) repeated measures ANOVA. A sensitivity power analysis conducted in GPower (version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007) showed that this analysis, with the current sample size and significance criterion, should have sufficient power ($\beta = .80$) to detect small to medium effect sizes ($f = .16$). Trials in which the participants answered after the question disappeared from the screen were excluded. This was the case for on average 3.2 trials for each participant and less than or equal to 10 out of all 340 trials for all participants.

Mauchly's test indicated that the assumption of sphericity was violated for the main effect of skill level, $\chi^2(9) = 33.12$, $p < .001$, and we therefore report the Greenhouse-Geisser corrected degrees of freedom for this effect, but not for the other effects. The results show no interaction between team and skill level, $F(4, 116) = .60$, $p = .67$, $\eta_p^2 = .02$. They do, however, show that the predicted score depends significantly on both the team of the player, $F(1, 29) = 9.64$, $p = .004$, $\eta_p^2 = .25$ and her average skill level, $F(2.60, 75.45) = 20.96$, $p < .001$, $\eta_p^2 = .42$ (see Figure 2). More specifically, participants overestimated the performance of members of their own team: they predicted the scores of their team to be higher ($M = 5.56$, $SE = .11$) than those of the other team ($M = 5.11$, $SE = .94$), an effect that did not correlate with individual participants' competitive attitude. In addition, they seem to have taken the average score of each player into account, as the tests of within-subjects contrasts showed a significant linear trend, $F(1, 29) = 45.28$, $p < .001$, $\eta_p^2 = .61$. The higher the player's skill level, the higher the predicted score. The same effect was found in separate analyses for ingroup, $F(1, 29) = 38.95$, $p < .001$, $\eta_p^2 = .57$ and outgroup, $F(1, 29) = 30.55$, $p < .001$, $\eta_p^2 = .51$, suggesting that participants individuate players in both teams. This is in line with the fact that there is no interaction between team and skill level, which suggests that there is no difference in individuation between ingroup and outgroup.

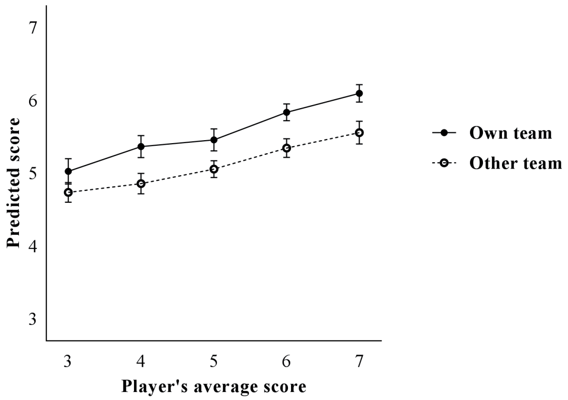


Figure 2. Predicted scores for players (mean ± SEM) with different skill levels from the own and the other team.

DISCUSSION

In the current study, we investigated the models underlying predictions about individual task performance, using a paradigm in which participants predicted the scores of bowling players from their own and another team. There are two main findings. First, our results show that participants distinguished between outgroup members in the same way they distinguished between ingroup members. This means that in a specific setting such as the one created here, people's person performance models of both ingroup and outgroup are based on the performance of individual players, allowing them to individuate members of both social groups. Second, participants overestimated the average performance of members of their own team, suggesting that the person performance model for each member of our own social group incorporate a positive bias, while members of the other social group are represented more veridically.

Of course, participants were not able to predict the performance of the players perfectly. As the slopes in Figure 2 show, the predictions for all players reflect regression to the mean, probably caused by the fact that when participants were unsure of their answers, they predicted a more average score of around 5. However, on average, estimates for the outgroup were more accurate than those for the ingroup, as the performance of ingroup members was overestimated.

Interestingly, these findings suggest that the bias causing ingroup overestimation is a robust one that persists even when one individuates members of all groups and when one has reliable online information about people's performance. In our controlled experimental setting participants were provided with clear information about performance and they could infer the accuracy of their predictions. Moreover, as we used a minimal group design, we did not group our participants according to existing social groups, but according to a random assignment to a group. It is interesting to see that even in this case, the difference in predictions about ingroup and outgroup members remains. It seems that the bias causing ingroup overestimation is a robust one that does not disappear when information about a person's actual performance is provided. This effect is unlikely to be caused by demand characteristics, as participants were simply instructed to predict the performance of different players and had no reason to overestimate members of their own team. Even if they would infer the aim of the study, they would be likely to not only overestimate their own team, but also underestimate the other team, which does not seem to be the case. Similarly, the findings unlikely to be caused by a simple increase in attention for the processing of ingroup information, as this should also increase the accuracy and amount of individuation of predictions about ingroup members, which is not the case. In this experiment, there is no indication that people's person performance model for the ingroup is more accurate or detailed than that for the outgroup. Moreover, the bias in the person performance model seems to be driven purely by the social group the observed person belongs to.

The observed individuation in both groups is in contrast with previous studies showing that we are better at recognizing faces of ingroup members (Bernstein et al., 2007; Van Bavel et al., 2012) and studies showing increased neural responses to actions of ingroup members (Molenberghs et al., 2013). As hardly any studies have directly investigated the person performance models underlying predictions when reliable information about individual task performance is available, it is difficult to compare circumstances in which individuation does and does not appear. One key difference between studies, however, might be the relative importance of ingroup and outgroup members. This importance may be driven by the social context. For instance, in certain competitive settings it is more important to gather information on the appearance or behaviour of outgroup members, potentially resulting in more individuation of the outgroup. This is in line with a study by Judd and Park (1988), who showed that under anticipated cooperation, memory for ingroup members was better than for outgroup members, but that anticipated competition increased memory for individual outgroup members to the level that there was no longer a difference between memory for ingroup and outgroup. In the current study, we also created a competitive setting, which may have caused participants to individuate in both groups. This would explain the contrast with the studies on face recognition, as those studies used more neutral settings in which the outgroup may have been perceived as less relevant (Bernstein et al., 2007; Van Bavel et al., 2012).

Another difference with previous studies may be the difficulty of the task or, in other words, the amount of information to be stored in the cognitive model. In this study, participants were presented with ten players for which information needed to be stored, which was less than the 40 faces that were shown in the studies by Bernstein et al. (2007) and Van Bavel et al. (2012). Therefore, it may have been rather easy for our participants to distinguish between individuals in both their own and the other team.

This potential role of the amount of information that needs to be represented may also explain the difference between this study and a study by Rubinstein, Jussim, and Stevens (2018). In that study, it was found that when people receive individuating information about two individuals from different racial groups, they are able to overcome the stereotype bias in their judgements of intelligence. Due to the experimental setting, the stereotype bias investigated in that study could be interpreted as similar to the ingroup bias we investigated here. Possibly, as participants only processed information about two individuals, they may have been able to judge their characteristics more accurately than when they would have when judging information about ten different individuals.

Importantly, our participants seem to have represented information about ingroup and outgroup members at similar levels of detail but still overestimated performance of the ingroup, which suggests that ingroup overestimation is not simply the result of having more detailed knowledge about a certain social group. This also implies that gaining more information about outgroup members does not necessarily reduce ingroup overestimation.

We think that our findings have potential relevance in the light of societal issues. Most importantly, they suggest that even if we believe that our predictions of another person's performance are based on objective information about their previous performance, they seem to be influenced by the fact that the person we are judging is part of one or another social group. If this happens in a controlled setting as the one we created here, the effect may be even stronger in more realistic settings, where less information about actual performance is available. Imagine, for example, a job interview in which the applicant has a different cultural background than the interviewers. If this person is indeed regarded as a member of an outgroup, then his or her skills may be judged less positively than those of another applicant who has exactly the same skills, but is regarded as a member of the ingroup. If so, then this may explain why employers tend to hire people that are culturally similar to them (e.g. Rivera, 2012) and why online resumes of people with a name associated with an ethnic minority group are viewed less often than those of people with a name associated with an ethnic majority group (Blommaert, Coenders, & Van Tubergen, 2013). Our findings suggest that such effects may persist even if employers carefully consider information about people from both ingroup and outgroup and thus individuate in both groups. This may be one of the reasons why it turns out to be so difficult for companies to diversify their workforce. If we indeed overestimate the future performance of people that are part of our own social group, we will give them more chances and are more likely to offer them a position. Similar problems may occur in other settings in which people are selected from a large number of applicants.

It seems of great societal relevance to further investigate if the ingroup overestimation that we have observed when reliable performance data for both ingroup and outgroup is provided also occurs in societally relevant scenarios, and if so, how people can be made to evaluate the performance of ingroup and outgroup members comparably. In any event, our research shows that stimulating outgroup individuation may be insufficient to overcome this bias when it occurs.

5

Predictions in action observation: the role of causal interpretation

ABSTRACT

Prior knowledge about the skill level of a person allows us to form expectations about the outcomes of this person's actions. In the study described in **chapter 2**, we found that this facilitates the processing of expected outcomes and suggested that this was driven by knowledge about the causal relation between the agent and the outcome. It is unknown, however, whether expectations are only based on causal knowledge about agents and their action outcomes (i.e. cause) or whether they can also be based on causal knowledge about something that enables the agent to cause a certain outcome (i.e. enabling condition). In order to increase our understanding of the role different types of causal knowledge in expectations of action outcomes, we here explore the latter situation. Participants viewed animated movies of people playing a bowling game in which the score was associated with different types of cues. In the first experiment, the colour of the bowling lane predicted either a high or a low score. The colour could be interpreted as an indication of the type of lane and thus, possibly, also of its quality, which is an enabling condition in the causal chain between the player's actions and the score. Participants were faster to answer questions in case these questions followed movies in which the score was as could be expected based on the colour of the lane. In two follow-up experiments, such an expectancy effect also occurred in situations in which the cue was not an enabling cause. Together, these findings suggest that cognitive models that aid processing are not only able to represent relations in which the cue is related to the causing agent itself but also to an enabling condition and, surprisingly, even to a seemingly causally unrelated part of the scene. We discuss these findings in the light of cognitive models representing causal or non-causal relations between cues and events.

INTRODUCTION

In many cases, what we have learned or experienced before may help us understand, predict and, if necessary, quickly respond to new events. Imagine, for instance, seeing different people trying to open a door: the first time you see someone trying but failing, you might be surprised that the door does not open. However, when a second person comes by and also tries to open it, you will probably have realized that the door is locked and not be as surprised anymore if this person also fails at opening it. Thus, prior knowledge, such as the knowledge about a door being locked and the causal relations between events, allows us to build up expectations of the outcomes of other people's actions.

It has been suggested that such expectations may actually facilitate the processing of events, when events are as expected. For instance, a variety of expected events have been found to be associated with decreased reaction times (e.g. O'Reilly et al., 2013; Van Elk et al., 2009) and brain activity (e.g. Summerfield et al., 2008; Vander Wyk et al., 2009) as compared to unexpected events.

Expectations arise in cognitive (or generative) models representing prior knowledge. As knowledge about causal relations is often a useful source of information on what to expect, it seems likely that such knowledge plays a key role in observers' cognitive models. This is in line with the idea that the degree to which two events are associated can partly be guided by knowledge about causal relations, as we have suggested previously (Heil et al., 2014). This suggestion was based on several findings. For instance, it was found that people can only be conditioned to blink their eyes after hearing a tone if they have previously learned that there is a relation between this tone and a puff of air to the eye (Clark & Squire, 1998). In addition, it was found that the blocking effect in conditioning (i.e. the effect that the association of an event A with event Y is prevented if A is presented together with another event B that has previously been associated with event Y) depends on by whether participants believe that A and B are either possible causes or possible effects of Y (Waldmann, 2000). Together, these findings suggest that if observers are aware of a causal relation between events, they are more likely to associate them and thus expect a second event to occur based on observation of the first event.

In the study described in **chapter 2**, we investigated the processing of action outcomes that could be predicted based on colour cues. These colour cues could either be interpreted as part of or unrelated to the causal chain of events. In the first of three experiments, participants watched animated movies of people playing a bowling game and reported either which player they saw or which score he obtained. The score could be predicted based on knowledge about the skill level of the player, as he was labelled as 'novice' or 'experienced' and usually scored in accordance with this level of expertise. Participants were faster to report which player they saw if the outcome (i.e. a high or a low score) was expected as compared to when it was not expected. In one of the follow-up experiments, participants watched similar movies, but now the score could be predicted based on a colour patch next to the player with the same predictive value as the skill level of the player in the first experiment. This time no effect of expectancy was found. Based on these findings, we suggested that in a naturalistic situation such as the one created in the experiment, expected events are processed more efficiently in case the relation between the cue (in this experiment the agent or the colour patch) and the subsequent event (the pins falling down) can be interpreted causally. This would mean that the expectancy effect in the first experiment was guided by the causal interpretation of the situation: knowing that players cause bowling pins to fall down may allow for predictions of the score based on knowledge about the players, whereas the lack of a causal interpretation may have prevented the formation of predictions in the second experiment.

In causal relations, there usually is a (human or non-human) agent causing a certain event to occur. However, it might depend on other factors whether an agent is actually able to cause the event. In the example described earlier, whether the door opens does not only depend on whether someone tries to open it, but also on whether it is locked or not. If we distinguish between causes and enabling conditions (as suggested by e.g. Cheng & Novick, 1991; Goldvarg & Johnson-Laird, 2001), a person's movements aimed at opening the door can be seen as the cause of the door opening and the door being unlocked as the enabling condition. In our previous study, the cue either served to identify the cause (i.e. the bowling player) or an object that was not part of the causal relation (i.e. the colour patch).

It is unknown, however, whether the expectations that are built up based on prior experiences and existing knowledge take cues about enabling conditions into account. For instance, independent of the specific movements of a bowling player, a certain type of bowling lane might cause a smoother movement of the ball and therefore a higher score. A cue indicating the type of bowling lane could therefore be used to build up an expectation of the score. Such an expectation might aid efficient processing and thus cause a shortening of reaction times.

Therefore, in this study, we aimed to investigate whether an expectancy effect indeed occurs in case expectations can be based on an enabling condition rather than on the cause itself. In the first experiment, we manipulated the colour of the bowling lane which, as suggested before, could be interpreted as an indication of the type of lane. There was a probabilistic relation between the colour and the score in the sense that one colour was associated with a low score and one with a high score. Therefore, participants might believe one lane is better than the other and as such, the colour may help them build up an expectation of the score based on the lane colour. In other words, they may interpret the type of lane as an enabling condition that partly determines whether the actions of the player cause a high or a low score. If such a relation is indeed represented in the participants' cognitive models, they are expected to be faster to answer questions if these questions are preceded by a score that is expected based on the lane colour.

This would suggest that cognitive models that facilitate processing are able to represent relations in which the cue is not related to the causing agent itself but also to another factor in the causal chain of events. In order to further investigate how robust such an effect would be, we follow up on this experiment with experiments in which the interpretation of the cue as an enabling condition in the causal chain is less obvious, such as when it is not the lane itself but only the side that is coloured or, as in the previous study, when a colour patch is shown next to the player.

The results of this study may help us understand how different types of causal information might be used to predict the outcomes of other people's actions and, as such, provide insight in the cognitive models underlying our expectations.

EXPERIMENT 1

Methods

Participants

Thirty-two healthy participants (23 female) between the ages of 18 and 33 (mean age 22.5) took part in this experiment. All participants were paid 10 euros or received course credits for their participation. The study was approved by the institution's local ethics committee and written informed consent was obtained from each participant.

Stimuli and design

For this experiment, 24 animated movies with a duration of 4500 milliseconds were created using Autodesk's 3ds Max 2014 and MotionBuilder 2014 (www.autodesk.com). Each movie showed a male bowler throwing a ball directed at ten pins standing upright at the end of a bowling lane. The avatar for the player was selected from Worldviz Vizard Complete Characters (www.worldviz.com/products/avatars/complete-characters). In order to avoid distraction, the player disappeared at 1300 ms after the start of the movie. The bowling ball rolled towards the end of the lane, either slightly left or slightly right of the centre, and knocked down 1, 2, 3, 6, 7, or 8 pins. The kinematics of both the action kinematics and the ball movement only varied with movement direction (i.e. left or right) and were not associated with a certain score.

In contrast, the colour of the bowling lane was associated with a certain score. For half of the participants, a blue lane was associated with a high score and a green lane was associated with a low score. For the other participants, this association was reversed. In 75% of all 432 trials, the score was as could be expected based on colour of the lane. More specifically, in 162 out of 216 trials showing a bowling lane in one colour, the score was low (i.e. 1, 2, or 3), and in 162 out of 216 trials showing a bowling lane in the other colour, the score was high (i.e. 6, 7, or 8). In the analysis, all these trials were labelled 'predicted' and compared to 'unpredicted' trials. Over all trials, the most frequent scores were 2 and 7, which appeared in 384 out of 432 trials. The other scores (i.e. 1, 3, 6, and 8) were included as fillers aimed to make the experiment more realistic by providing variability in scores.

Procedure

Upon arrival, participants were seated in front of a computer on which the experiment was presented. Stimuli and instructions were displayed using Presentation software (version 18.1, www.neurobs.com). Participants were asked to follow the instructions on the screen, which also explicitly informed them that the bowling lane could be either green or blue and that this colour was associated with a certain score. Before starting the main task, participants performed four practice trials in which they were reminded about this association between the colour and the score.

In all trials, participants watched an animated movie of a bowling event and answered a question about it afterwards (see Figure 1). This question either concerned the colour of the lane ('Was the bowling lane blue or green?') or the score ('Was the score high or low?'). Participants were unaware which one of these questions they would be asked after each movie, as they appeared in random order. To answer the question, participants could choose between two answer options, presented underneath the question on the screen. Participants were instructed to answer as quickly as possible by pressing either the left or the right button on a button box. The order of the answer options was randomized to prevent motor preparation. The question and the answer options stayed on screen for

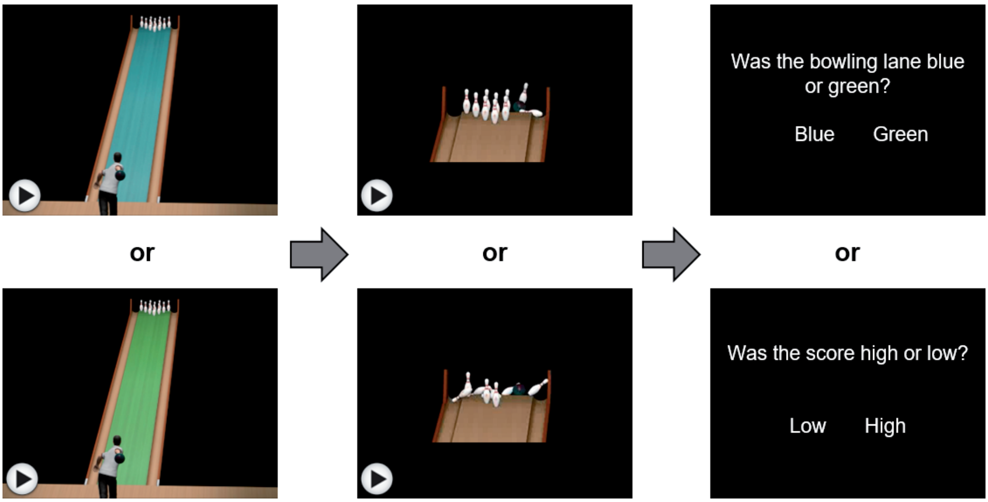


Figure 1. Schematic representation of the procedure in Experiment 1. In each trial, participants saw an animated movie of a person throwing a ball on either a blue or a green bowling lane and obtaining either a low or a high score. Each movie was followed by a question about the colour of the lane or the score.

2000 milliseconds or until the question was answered. Participants received feedback on the accuracy of their answer for the practice trials, but not for the other trials. After each trial, a fixation cross was presented for 700 milliseconds.

Reaction time analysis

We anticipated that answering a question that followed an unpredicted score would take more time than answering a question that followed a predicted score. Reaction times were analysed using a 2 (prediction: predicted vs. unpredicted) × 2 (question: colour vs. score) repeated measures ANOVA. In this analysis, only trials in which the score was 2 or 7 were considered, since the other scores (i.e. 1, 3, 6, or 8) were only included as fillers and appeared very infrequently. Reaction times to questions that were answered incorrectly were excluded from the analysis.

Results and discussion

The analysis showed no significant interaction between prediction and question, $F(1, 31) = 1.29$, $p = .26$, $\eta_p^2 = .04$. However, it did show significant main effects of both prediction, $F(1, 31) = 4.85$, $p = .04$, $\eta_p^2 = .14$, and question, $F(1, 31) = 15.67$, $p < .01$, $\eta_p^2 = .34$. On average, the reaction time was 9 ms longer for questions following an unpredicted score than for questions following a predicted score and 51 ms longer for the colour question than for the score question (see Figure 2).

5

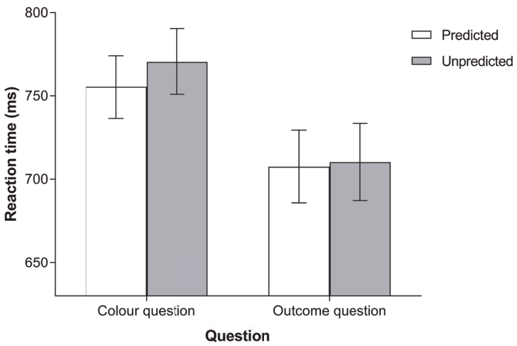


Figure 2. Reaction times (mean ± SEM) for the colour question and the outcome question in Experiment 1, with separate bars for predicted (white bars) and unpredicted outcomes (grey bars).

This means that whether or not a score is as predicted based on knowledge about the bowling lane determines at which speed a question will be answered. This in line with the idea that predictions play a key role in the processing of new events and with the idea that such predictions can be based on a relation between the bowling lane and the score. The effect of question indicates that participants needed more time to answer the question about the colour than the question about the outcome, potentially because of the specific phrasing or characteristics of these questions. Furthermore, the lack of an interaction effect between question and prediction suggests that the expectation effect does not differ between the questions.

In sum, the results suggest that participants used the cue related to the enabling condition (i.e. the colour indicating the type of bowling lane) in order to build up expectations of the score. This would mean that their cognitive models are not only able to represent relations in which the cue is not related to the agent itself, but also relations in which it is related to an enabling condition that is part of the causal chain between the agent and the outcome. Although it seems that participants have interpreted this situation causally, we cannot be sure that the possibility to interpret the relation between the colour of the lane and the score causally was indeed crucial. Therefore, we set up a follow-up experiment in which there is no obvious causal interpretation. In this experiment, the colour of the side of the lane will be manipulated and there will be a link between this colour and the score with the same predictive value as in the current experiment. In this case, as the bowling ball does not touch the side of the lane, there is no obvious causal relation between its colour and the score. We would therefore expect no expectancy effect that facilitates processing.

EXPERIMENT 2

Methods

Participants

Thirty-two healthy participants (22 female) between the ages of 18 and 34 (mean age 22.5) took part in this experiment. An additional two participants completed the experiment but were excluded from the analysis because their accuracy for one of the questions was below 80%. Again, all participants were paid 10 euros or received course credits for their participation. The study was approved by the institution's local ethics committee and written informed consent was obtained from each participant.

Stimuli and design

The design of this experiment was similar to that of the first experiment. The only difference was that we now used animated movies in which it was the side rather than the central part of the bowling lane that had a certain colour. The associations of the colours with a certain score and the trial distributions remained exactly the same.

Procedure

In this experiment, we followed the almost same procedure as in the first experiment (see Figure 3). Only this time, participants were informed that the side of the bowling lane could be either green or blue and that this colour was associated with a certain score. The phrasing of the colour question was adjusted accordingly (i.e. 'Was the side of the bowling lane blue or green?').

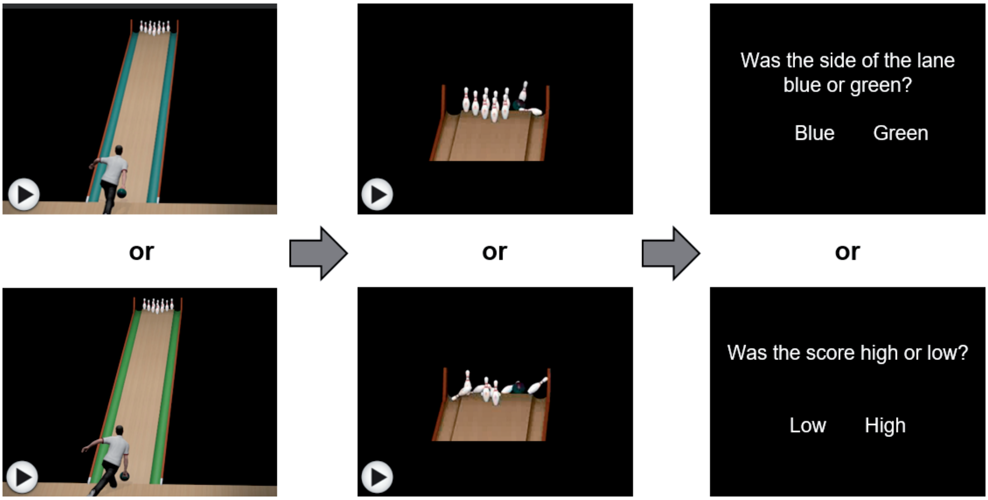


Figure 3. Schematic representation of the procedure in Experiment 2. In each trial, participants saw an animated movie of a person throwing a ball on a bowling lane with either blue or green sides and obtaining either a low or a high score. Each movie was followed by a question about the colour of the lane or the score.

Reaction time analysis

As in Experiment 1, we analysed reaction times using a 2 (prediction: predicted vs. unpredicted) × 2 (question: colour vs. score) repeated measures ANOVA, considering only trials in which the score was 2 or 7 and the question was answered correctly.

Results and discussion

As in the previous experiment, the interaction between prediction and question was not significant, $F(1, 31) = .11$, $p = .74$, $\eta_p^2 < .01$, but there was a significant main effect of prediction, $F(1, 31) = 5.39$, $p = .03$, $\eta_p^2 = .15$. The reaction time to questions following an unpredicted score was on average 13 ms longer than the reaction time to questions following a predicted score. Also, there was a significant main effect of question, $F(1, 31) = 9.89$, $p < .01$, $\eta_p^2 = .24$. The reaction time to the colour question was on average 44 ms longer than for the outcome question (see Figure 4).

These findings are similar to those of the previous experiment. It is possible that participants have interpreted the relation between the colour of the side of the lane and the score as a causal one. Just like the colour of the lane in the first experiment, the colour of the side of the lane may have been interpreted as an indication of the type or quality of the lane. As such, it may have caused participants to interpret it as a factor that influences whether the score will be high or low. Therefore, in another follow-up experiment, we manipulated the stimuli in such a way that the cue cannot be interpreted as part of the causal chain of events by changing the colour of a patch shown next to the player.

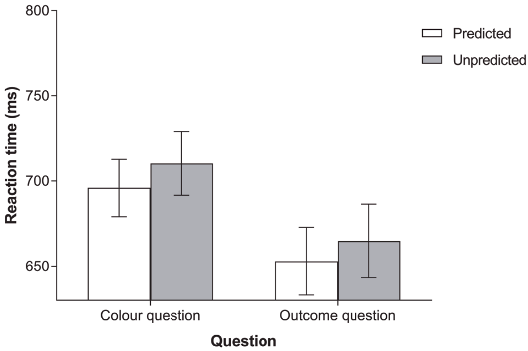


Figure 4. Reaction times (mean ± SEM) for the colour question and the outcome question in Experiment 2, with separate bars for predicted (white bars) and unpredicted outcomes (grey bars).

EXPERIMENT 3

Methods

Participants

Thirty-two healthy participants (23 female) between the ages of 18 and 33 (mean age 22.7) took part in this experiment. An additional two participants completed the experiment but were excluded from the analysis because their accuracy for one of the questions was below 80%. As in the previous experiments, all participants were paid 10 euros or received course credits for their participation. The study was approved by the institution's local ethics committee and written informed consent was obtained from each participant.

Stimuli and design

The design of this experiment was similar to that of the other two experiments. The animated movies used in this experiment differed in the sense that they now showed a bowling lane in a neutral colour, with a square patch in a certain colour presented next to the player. This patch was either blue or green and this colour was associated with the score, with exactly the same regularity as the colour of (the side of) the bowling lane in the previous experiments.

Procedure

In this experiment, we followed almost the same procedure as in the previous experiments (see Figure 5). Participants were informed that the colour patch could be either green or blue and that this colour was associated with a certain score. Again, the phrasing of the colour question was adjusted accordingly (i.e. 'Was the patch blue or green?').

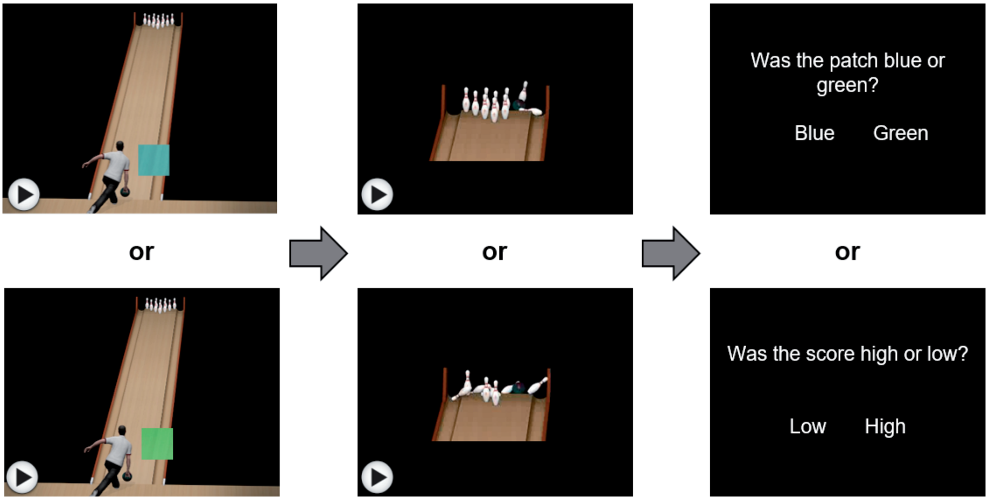


Figure 5. Schematic representation of the procedure in Experiment 3. In each trial, participants saw an animated movie of a person throwing a ball on a bowling lane and obtaining either a low or a high score. Next to the person, either a blue or a green patch was shown. Each movie was followed by a question about the colour of the lane or the score.

5

Reaction time analysis

As in Experiments 1 and 2, we analysed reaction times using a 2 (prediction: predicted vs. unpredicted) × 2 (question: colour vs. score) repeated measures ANOVA, considering only trials in which the score was 2 or 7 and the question was answered correctly.

Results and discussion

The analysis showed no significant interaction between prediction and question, $F(1, 31) = .03$, $p = .87$, $\eta_p^2 < .01$. This is in line with the previous experiments, as are the significant main effects of prediction, $F(1, 31) = 11.65$, $p < .01$, $\eta_p^2 = .27$, and question, $F(1, 31) = 11.67$, $p < .01$, $\eta_p^2 = .27$. The reaction time was on average 22 ms longer for questions following an unpredicted score compared to questions following a predicted score and 49 ms longer for the colour question than for the score question (see Figure 6).

Again, these findings are similar to those in the first two experiments. As in those experiments, it seems that predictions are built up based on the association between the colour and the score and that these predictions speed up processing of expected events. We hypothesized that the lack of a causal interpretation in this situation (i.e. there is no plausible explanation of why a colour patch would be part of the causal chain and thus influence the score) would prevent participants from building a cognitive model in which the score can be predicted based on the colour. If this would indeed be the case, then we would not expect a difference in reaction times between the predicted and the unpredicted events.

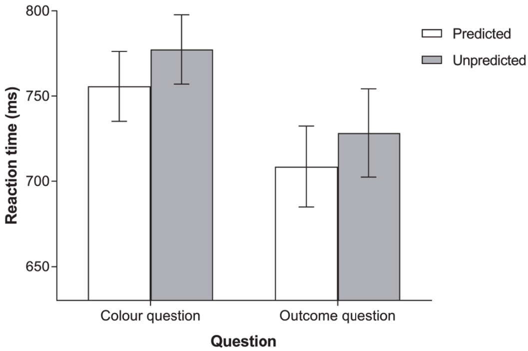


Figure 6. Reaction times (mean ± SEM) for the colour question and the outcome question in Experiment 3, with separate bars for predicted (white bars) and unpredicted outcomes (grey bars).

In order to determine if there was a difference in the expectancy effect between the three experiments, we analysed the data from all three experiments in one ANOVA, with prediction and question as within-subject factors and experiment as a between-subject factor. This analysis shows that there is no interaction between prediction and experiment, $F(1, 93) = 1.42$, $p = .25$, $\eta_p^2 = .03$. This means that there is no evidence that the expectancy effect differs between experiments.

DISCUSSION

The aim of this study was to investigate how different types of causal information might be used to predict the outcomes of other people's actions. We assumed that causal knowledge plays a key role in building up expectations about action outcomes and therefore, that whether or not people expect a certain action outcome based on a predictive cue would depend on whether the cue was part of the causal chain between the agent and the outcome. Surprisingly, however, expectancy effects were found in all three experiments, including at least one in which there was no obvious causal relation between the cue and the action outcome. This suggests that cognitive models that aid processing are not only able to represent relations in which the cue is related to the causing agent itself but also relations in which the cue is related to an enabling condition or even to a seemingly unrelated part of the scene.

In the first experiment, we examined whether participants would use knowledge about an enabling condition in the causal chain of events, such as the type of bowling lane in the first experiment, to build up expectations about the action outcome. This indeed seemed to be the case: participants seem to have used the colour cue to predict the score. Contrary to our expectations, however, the results of the two follow-up experiments suggest that participants also used the colour cue to predict the score when a causal interpretation was unlikely. In the second experiment, the colour cue was part of the side of the bowling lane. One could argue that in this case, the colour also used as a way to identify the type of lane and therefore could still be part of the causal chain of events. However, there was no reason why the colour patch in the third experiment could be seen as part of this causal chain, as it was not related to the player, to the lane, the ball or the pins.

Interestingly, the expectancy effects we found in all three experiments differ from those found in the study described in **chapter 2**. In that study, as described in the introduction, participants first saw different players who obtained different average scores. We not only found an expectancy effect in the sense that responses to questions following unexpected events were slowed down as compared to those following expected events, but we also found that this expectancy effect specifically occurred when participants reported which player they saw. When participants reported which score they saw, there was no expectancy effect. Our interpretation of this difference between questions was that the cognitive model from which predictions arise represents player and score in a hierarchical relation. Information about the score is represented at the level below that at which information about the player is represented. Our assumptions for this model were derived from the predictive processing framework, which suggests that the brain is primarily focused on processing the difference between predictions and actual sensory input. The predictions are sent from higher cognitive levels down to lower levels, whereas prediction errors that result from the comparison are sent back up in order to improve the predictions (Clark, 2013b; Friston, 2010). In terms of the previous study, in case an event is not as expected, this causes a prediction error at the level of score, which is sent to the level above, where the cognitive model needs to be updated in order to resolve the prediction error. We then suggested that because this requires additional processing at the higher level, reporting of the cause (i.e. the player) is slowed down. As a result, an unexpected event might specifically slow down reporting of the player as in the agent question, but not influence reporting of the score in a similar manner.

Assuming that this interpretation is correct, one might expect a similar difference between questions in the current study. However, in this study, the expectation effects show up in both questions: participants are slower to respond to a question that follows an unexpected score, independent of whether this question is about the colour or the score. Potentially, this could mean that none of the cues is actually interpreted as part of the causal chain. The colour of the bowling lane (or its side) may not have been interpreted as an indication of the type of lane and thus of its potential to enable a high score. In this case, the expectancy effect would have been caused by a non-hierarchical, non-causal association between a colour and a score. In our previous study, we suggested that links between levels in the cognitive model have a causal interpretation, whereas links within one level do not need such causal interpretation. The idea that links between levels have a causal interpretation is in line with the view that, according to predictive processing, top-down predictions infer sensory data from high-level causes and that, therefore, the causal structure of the world is reflected in backward connections (Friston, 2005). Our current findings might then be an indication that the cognitive model that observers use does not represent information related to the cue at a higher level than information related to the action outcome. Instead, the different types of information might be represented at one level and purely associative links that have a non-causal interpretation.

Such a difference between models with hierarchical, causal links on the one hand and non-hierarchical, non-causal links on the other hand might also be able to explain another difference between this and our previous study. In real-world situations, it seems likely that causal associations are usually more informative: knowledge about the mechanism causing a certain event often helps building up correct expectations about the occurrence of this event. This may also explain why, when observing the actions of other people, we do not simply interpret these actions as associated with a certain outcome, but interpret these actions as a means to reach a certain goal (Csibra & Gergely, 2007). Based on this idea, it is conceivable that non-causal associations lead to weaker predictions than causal ones (Tversky & Kahneman, 1980). Causal associations, on the other hand, may be more selective. It has been found, for instance, that if one cue is assumed to cause a certain outcome, the other cue is less likely to be interpreted as a cause of the same outcome (Pineño, Denniston, Beckers, Matute, & Miller, 2005). Potentially, there is a balance between strength and specificity: whereas causal associations are rather specific but strong predictors of an outcome, non-causal associations are broader, but weaker predictors of this outcome. Assuming that the predictions in two of the experiments of the previous study were based on causal associations, while those in the current study are based on non-causal associations, this might explain why the expectancy effect is larger in the previous study than in the current study. In the two experiments in the previous study in which an expectancy effect was found, the average difference in reaction time between expected and unexpected trials for the question about the agent or colour was around 37 ms. However, in the same question in the current study, the difference was only around 17 ms. Such an interpretation would be in line with inferential reasoning accounts of human contingency learning, that assume that top-down inferential processes play a role in the process of learning that two events are related (see Pineño & Miller, 2007 for an overview of different accounts).

The last experiment in this study is actually very similar to the second experiment in the previous study. In both experiments, participants could base their predictions about the score on a colour patch next to the bowling player and the probabilistic relation between the cue and the score were the same. However, whereas we found no expectancy effect in the first study, we did find it in this study. Potentially, this is the result of a lower power in the first study, as that study involved fewer trials and participants. Together with the difference in reaction time effects, this might imply that there is no qualitative difference, but primarily a quantitative difference between different types of associations.

It would be interesting for future research to further investigate if there are indeed be different types of links between representations in cognitive models, which differ in terms of their strength or ease of learning. Furthermore, assuming that participants in this study did not build a cognitive model in which the cue was causally linked to the outcome, we cannot be sure why this is the case. Given the distinction between causes and enabling conditions, it is possible that a relation is only encoded as causal when the predictive cue is linked to the cause itself rather than to an enabling condition. There are, however, also alternative explanations.

One alternative explanation is that participants were not aware of the causal mechanism, as this was not explicitly stated in the instructions. In the first experiment, participants were told that the colour of the bowling lane was associated with a certain score, but not that one bowling lane was of better quality than the other. Although causal interpretation is indeed likely to depend on an understanding of the causal mechanism (Shultz et al., 1986), the findings of our previous study suggest that this interpretation can also arise without explicit instruction. In the last experiment of that study, we did find an expectancy effect even though there were no explicit instructions about the player being a 'novice' or 'experienced' player, but instead there were implicit instructions stating that the colour of the player's shirt was associated with a certain score.

Another explanation might be that in order for a causal association to be made, the cue needs to be related to the agent causing the action. It would be interesting to investigate whether this is indeed the case and whether this agent needs to be a person, or whether it can also be a non-human (or even inanimate) agent.

In sum, our findings suggest that people build up expectations about action outcomes based on a cue that can be used to identify the cause of the action or an enabling condition in the causal chain of events. Surprisingly, expectations can also be based on a cue that is not part of the causal chain of events. We suggest that this may indicate that different types of links between a cue and an outcome differ in terms of the strength of predictions arising from an observer's cognitive model. As such, these findings provide insights in the cognitive models underlying our expectations and open up interesting questions for future research on this topic.

6

Summary and discussion

The aim of this thesis was to investigate whether the social-cognitive processing involved in understanding and predicting other people's actions can be explained within the predictive processing framework. To that end, four studies were conducted in which hypotheses about the potential nature of the generative models underlying predictions were formulated and tested. In all studies reported in this thesis, a newly-developed bowling paradigm was used. In this paradigm, participants viewed animated movies of people playing a bowling game. Based on different types of information, such as knowledge about the player, they could build up expectations of the most likely obtained score. **Chapter 2** described a study in which we investigated whether participants use knowledge about the skill level of an individual player in the processing of the player's actions. In the main experiment, participants saw either novice or experienced players obtaining a high or a low score and answered questions about the player or the score. We assumed that unexpected action outcomes would result in prediction errors, which would be reflected in longer reaction times. We found that a predictive processing model with a hierarchical structure and causal relations between its levels was able to explain the pattern of reaction times in a series of three experiments. In the fMRI study reported in **chapter 3,** participants performed the same task as in chapter 2. We found that unexpected action outcomes resulted in additional neural activity that seems to reflect explaining away of prediction errors. This activity showed overlap with that in a theory-of-mind task, suggesting that part of the prediction errors is explained away in brain areas typically associated with social-cognitive processes. **Chapter 4** reported on a study in which we explored whether the generative models of another person's task performance (i.e. person performance models) differ between ingroup and outgroup. Participants saw members of their own and another team playing a bowling game and predicted which score each player would obtain. The results suggest that although they distinguished between different members of both teams, they still overestimated the performance of members of their own social group. **Chapter 5** described a study focusing on how knowledge about different types of causal factors in the environment could be represented in a generative model (in that chapter referred to as cognitive model). Participants viewed movies of a person playing a bowling game. In these movies, there was a predictive cue that was related to factors that we assumed could (in the first experiment) or could not (in the subsequent experiments) be interpreted as enabling factors in the causal chain of events. The results suggest, however, that none of the situations may have been interpreted causally, but that participants were still able to predict the score based on the cues. We speculate that this may be an indication that non-causal factors may be incorporated in the generative model, be it less easily or with less predictive power. In this chapter, the findings will be discussed in terms of the potential nature of generative models underlying our predictions of other people's actions.

SOCIAL KNOWLEDGE IN GENERATIVE MODELS

People are usually quick to understand what another person is doing, even if this understanding depends on knowledge about the social context or characteristics of the person performing the action. Assuming that this can be understood within the predictive processing framework, such knowledge needs to be represented in a generative model. This generative model can then provide predictions that are tested against input received from more low-level areas in the brain. The findings described in this thesis are consistent with the idea that people make predictions based on a generative model that incorporates information about the person performing the action. Outcomes that are unexpected given the skill level of the agent seem to require additional processing as compared to expected outcomes, in line with the idea that prediction errors need to be explained away. This additional processing is reflected in longer reaction times (**chapter 2**) and increased neural activity (**chapter 3**).

In addition, our findings suggest that information about the social group a person belongs to also influences predictions of action outcomes. In the experiment presented in **chapter 4**, predicted outcomes for ingroup members were overall higher than for outgroup members. Ingroup members were expected to score better than outgroup members even though knowledge about individual performance was available and incorporated in the observers' generative models, which suggests that information about social groups may not simply be used to fill in missing information about individuals. Instead, information about individuals and their social groups may both have a separate influence on the way in which generative models represent another person's expected performance. This finding is especially relevant to societal issues, as it seems to suggest that more knowledge about the performance of individuals from another social group is not always enough to overcome ingroup bias.

HIERARCHICAL GENERATIVE MODELS

An important assumption in the predictive processing framework is that of hierarchical generative models: predictions are produced by higher levels in order to inform lower levels, whereas the effect of prediction errors propagates upwards. The findings reported in **chapter 2** are in line with the idea that prediction errors arise at one level, but are explained away at a higher level. In the computational model presented in that chapter, the higher level represents information about the agent while the lower level represents information about the outcome of this agent's action. Others have suggested other divisions between levels that are relevant for action understanding. For instance, Kilner (2011) focuses on kinematic, motor, goal and intention levels, while Spunt, Satpute and Lieberman (2011) distinguish between the how, what and why of an action, and Hamilton and Grafton (2006) describe levels representing movements, actions, immediate goals and task goals. Although the precise distribution of information over levels may depend on the task at hand and type of information processed, all seem to agree on the idea that levels are ordered according to their abstractness and, related to that, the time they require (Van Overwalle & Baetens, 2009). This is also the case in the model presented in **chapter 2**, where the skill level of the agent exists over a longer period of time and is more abstract than the action outcome.

We assumed that prediction errors arising at the level of the action outcome would be explained away at the level of the agent. In line with this assumption, the increased neural activity found the study described in **chapter 3** was found in several areas including the temporoparietal junction, an area typically associated with thinking about other people's thoughts and beliefs (Saxe & Kanwisher, 2003). Although the paradigm used in this thesis does not require considering other people's mental states, it does require considering their personal characteristics (i.e. their skill level). This finding suggests that prediction errors do not only arise in areas involved in rather low-level areas, such as those involved in visual perception, but also in areas involved in higher, social-cognitive processes, in line with the idea that predictive processing provides a unifying framework for brain functioning (Clark, 2013b; Friston, 2005, 2010; Hohwy, 2013).

IMPROVING GENERATIVE MODELS

The predictive processing framework essentially assumes that processing is facilitated by prior experiences and previously acquired knowledge, as represented in generative models. In case these generative models are incorrect or incomplete, the predictions that they provide will result in prediction errors. Learning occurs when these prediction errors are used to improve the generative models.

This idea that learning is guided by prediction errors is in line with the Rescorla-Wagner model, a model that was developed to explain in which situations classical conditioning occurs (Rescorla & Wagner, 1972). It assumes that learning takes place not simply because two events co-occur, but because this co-occurrence is unexpected and thus results in a prediction error. As such, this type of learning is more advanced than simple co-occurrence types of learning, that are often (although possibly incorrectly, see Keysers & Gazzola, 2014) considered Hebbian. Learning according to the principles of the Rescorla-Wagner model is consistent with the predictive processing framework (Den Ouden et al., 2008). As these principles are central to associative accounts, such as the associative sequence learning model (Cooper et al., 2013) and related Hebbian models (Keysers & Gazzola, 2014), these accounts can also be incorporated in the framework (Campbell & Cunnington, 2017). This would imply that the generative models responsible for predictions are, at least in part, developed and improved by the associative learning that occurs when prediction errors arise.

These processes are likely to occur at all levels of the hierarchy involved in the processing of other people's actions. At the level of processing that occurs in areas that are considered part of the mirror neuron system, this integrative view would be in line with previous suggestions that mirror neurons acquire their mirroring properties by associative learning (e.g. Cook et al., 2014; Heyes, 2010) and that, once they have developed these properties, their activity can be explained by predictive principles (Csibra, 2008; Kilner, Friston, & Frith, 2007a; Kilner et al., 2007b; Miall, 2003).

THE ROLE OF CAUSALITY IN GENERATIVE MODELS

Although associative learning seems to play a key role in the learning from experience, this type of learning may not paint the full picture. The findings from the studies described in this thesis suggest that learning may not only be guided by associative principles, but also by knowledge about the causal structure of the world. In **chapter 2**, three different experiments were described, of which the first two are most relevant to illustrate the role of causal knowledge. In the first experiment, participants could base their predictions of the outcomes of bowling actions on knowledge about the person performing the action, while in the second experiment, these could be based on a simple colour cue. Although both cues (i.e. the person in the first experiment and the colour in the second experiment) has the same predictive value in terms of contiguity and contingency, only the agent-related cue seemed to become represented in the participants' generative models and thus used to predict the score. We argued that knowing that agents, but not colours, can cause the bowling pins to fall down, created the distinction between the results of the two experiments. This is in line with our previous suggestion that higher-level cognitive processes, such as causal reasoning, can guide the formation of appropriate associations (Heil et al., 2014) and therefore, that the learning mechanism as proposed by associative sequence learning or Hebbian accounts may not be sufficient to explain how generative models are improved based on experience. This would be in agreement with previous studies showing that whether an association is learned or not depends on the causal interpretation of the events involved (e.g. Clark & Squire, 1998; Waldmann, 2000). The influence of whether or not a situation is perceived as causal or not also explains other findings, such as those showing that young children are more likely to predict the movement of an object based on the position of a causal agent rather than that of an inert object (Saxe et al., 2007) and that the remembered speed of a movement depends on the effect it seems to have caused in Michotte's (1963) launching effect paradigm (Kerzel et al., 2000).

This does not mean, however, that it is impossible to build up expectations about an event based on a cue that is not causally related to this event. For instance, previous studies have shown evidence for prediction errors even in paradigms in which there was no causal relation between the cue and the event (e.g. Den Ouden et al., 2010; Kok et al., 2012). In addition, in **chapter 5** of this thesis, a study was reported in which different types of cues were associated with high or low scores in a bowling game. Although it seems that participants may not have interpreted these cues as causal (or enabling) factors, they were slower to respond to questions after they saw scores that were unlikely given the cue. This would mean that they did incorporate the cues in their generative model and thus predicted the occurrence of specific outcomes.

Although more research on the role of causal knowledge in the generative models people use to process other people's actions is required, it seems likely that the extent to which associations are formed depends, at least in part, on people's knowledge of the causal structure of the world.

FUTURE OUTLOOK

In this thesis, I have taken the predictive processing framework as a starting point to explore how people process the actions of others. Inspired by the framework, I hypothesized that priors represented in generative models would reflect knowledge about the person performing the action and the causal structure of the world. Although the findings presented and discussed here are consistent with this view and therefore show the potential of the predictive processing framework, they do not directly test it. Before it will be possible to test the framework and compare it to other accounts, its exact characteristics need to be specified further. This not only goes for the predictive processing framework, but also for potential alternative frameworks. At this point, there do not seem to be any clearly defined alternatives that cannot be explained within the predictive processing framework. This need for a more clear definition is particularly important for research on abstract social-cognitive processes such as action understanding. In the experiments discussed in this thesis, the actions processed are relatively straightforward and although participants needed to consider personal characteristics of the players, they did not necessarily have to consider their mental states. The situation complicates quickly if people need to consider an increasing amount and abstractness of information about the people they observe. It is not obvious, for instance, how abstract concepts such as intentions are represented in the brain and how they give rise to predictions specific enough to be tested against sensory input. It is also not clear whether information that is assumed to be processed at one level of the hierarchy is necessarily also processed in one area of the brain. It would be interesting to investigate how the hierarchies suggested in computational models such as the one presented in **chapter 2** translate to neural systems.

Furthermore, future work may help improve our understanding of the role of causality in action processing in general and the way in which it can be understood within the predictive processing framework specifically. In **chapter 5**, for instance, we speculated about the difference between causal and non-causal relations in generative models and in **chapter 2** we suggested that links between levels in the hierarchy have a causal interpretation. Future research is needed to further investigate these ideas. Further insights may be gained from developmental studies, as these could potentially illuminate the role of causal knowledge in the way in which generative models develop. For instance, infants may not have the rich generative models required to understand the causality of a situation, so one could wonder how they come to acquire this understanding in the first place. Interestingly, however, it seems that even rats possess a basic understanding of causality (Blaisdell, Sawa, Leising, & Waldmann, 2006) and some have suggested that children are born with innate assumptions about causality (Gopnik et al., 2004; Spelke, Breinlinger, Macomber, & Jacobson, 1992). More research on the way in which learning to associate two events depends on knowledge about the causal structure of the situation is key to our understanding of the generative models underlying our predictions and, as such, to our understanding of how predictive processing could support action understanding.

Another direction that would be relevant for future research concerns the view that predictive processing provides a unifying framework of brain functioning. It has been proposed to explain a wide range of cognitive processes and alterations in these processes that may give rise to disorders such as autism spectrum disorders (e.g. Palmer, Seth, & Hohwy, 2015; Pellicano & Burr, 2012) and schizophrenia (e.g. Fletcher & Frith, 2009). It would be interesting to see how far the explanatory power of the framework reaches. We may distinguish to qualitatively different approaches for doing this: (1) one could start at the bottom of the hierarchy by investigating whether specific neural responses are consistent with predictive processing (e.g. Bastos et al., 2012; Kok, Bains, van Mourik, Norris, & de Lange, 2016) and move up from there or (2) start at the top of the hierarchy by investigating whether cognitive processes can potentially be explained by predictive processing (e.g. Koster-Hale & Saxe, 2013) and move down from there. In this thesis, I have taken a perspective that can inform the latter approach. As suggested earlier, this approach faces the challenge of coming up with critical tests of the assumptions made. The bottom-up perspective, on the other hand, faces the challenge of scaling up from lower to higher cognitive levels. If we want to figure out if predictive processing can indeed explain processing in the entire cognitive hierarchy, we will need research focusing on both directions and thus addressing both challenges.

SOCIETAL RELEVANCE

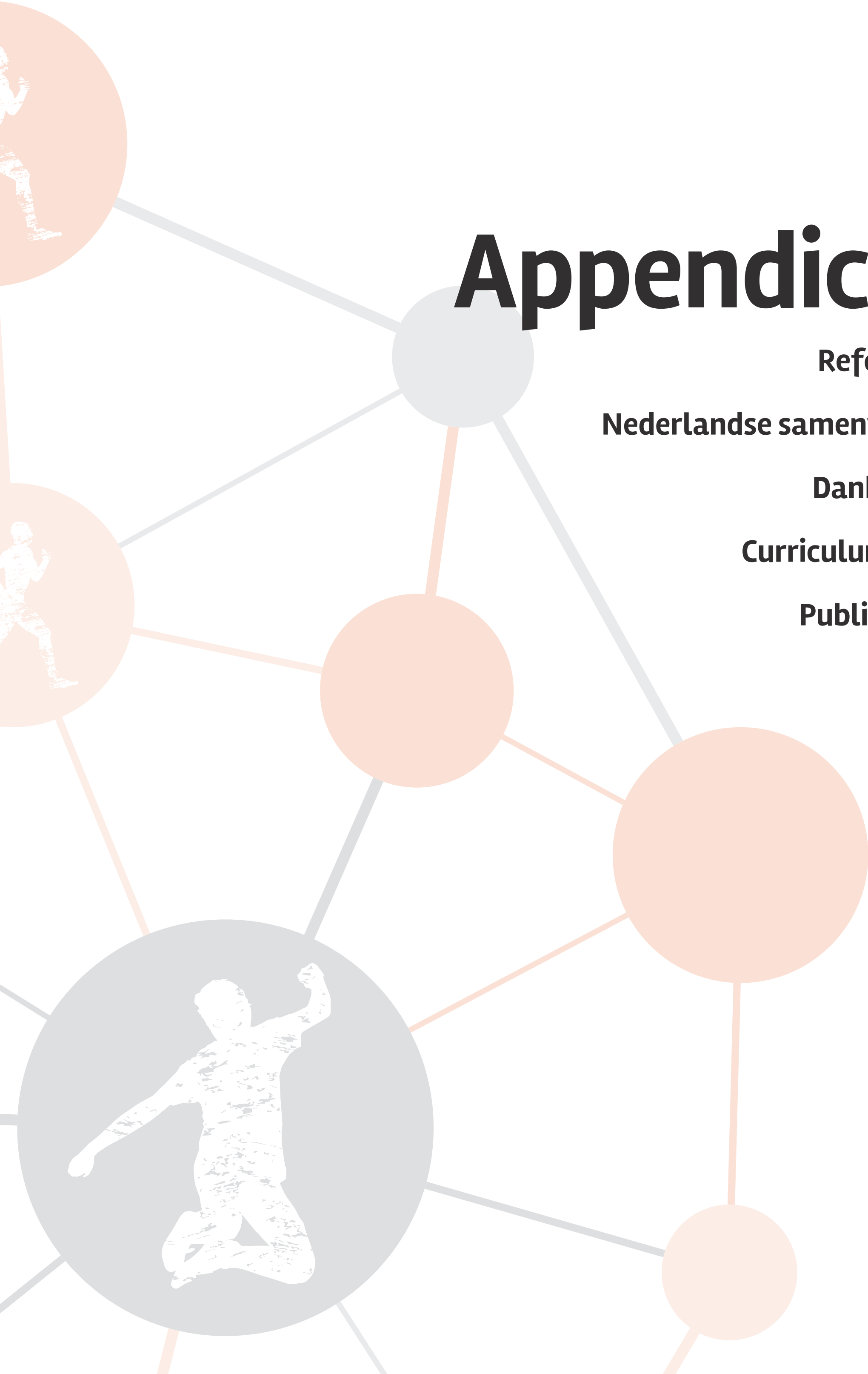
A better understanding of the way in which people process and predict the actions of others, may also provide insights of relevance for a range of societal issues. The view that brains are essentially prediction machines that learn from prediction errors may, for instance, improve our understanding of the way in which our judgements about others are influenced by stereotypes. The idea that information about social groups, including stereotypes, influences our judgements and even perception of others is generally accepted. However, understanding that this information is represented in generative models that provide us with predictions of people's actions and performance offers a clearer picture and may eventually even pave the way for better methods to overcome the negative impact of stereotyping. Such methods could help increase equality and avoid the automatic negative evaluations of people who act in a way that does not match stereotype images (Flannigan, Miles, Quadflieg, & Macrae, 2013).

Furthermore, the view suggested in this thesis may shed light on educational methods in two ways. First, if prediction errors are indeed essential for learning or, in terms of predictive processing, the improvement of generative models, then it seems important to create learning materials and methods that cause prediction errors. According to the learning progress hypothesis, the experience of learning progress (i.e. a reduction in prediction errors) can be intrinsically rewarding. It therefore leads to curiosity, which then fosters learning. This results in a positive feedback loop between curiosity and learning (Oudeyer, Gottlieb, & Lopes, 2016). In education, it then seems key to get children to generate predictions in order for them to experience learning progress. Second, the idea that learning that two events are related is more effective in case there is a clear causal relation between the events, highlights the importance of understanding the mechanisms behind their occurrence.

CONCLUDING REMARKS

The aim of this thesis was to investigate how people process other people's actions and the outcomes of these actions. The research presented here points towards an integrative view. It suggests that, in principle, all aspects investigated here can be understood within the predictive processing framework. Our findings were in line with the idea that action outcomes that are unexpected given a person's characteristics result in prediction errors and that these prediction errors are, at least in part, explained away in brain areas typically associated with social-cognitive processes. Furthermore, they suggest that predictions about action outcomes arise in generative models that have a hierarchical structure and incorporate knowledge about whether or not factors are causally related to the action outcome. As proposed earlier, the way in which generative models are developed and improved may largely follow the principles suggested by associative accounts of learning, complemented with principles that take causal knowledge into account. Finally, they suggest that the generative models that aid the processing of actions and their outcomes do not only represent knowledge about individuals, but also about the social group they belong to, as members of one's own social group seem to be represented more positively than members of another social group. Together, these findings suggest that the predictive processing framework cannot only explain low-level processes, but also more abstract social-cognitive processes such as those involved in action processing. However, future research would need to further explicate different aspects of the framework in order to compare it to alternative accounts and determine whether it would indeed be able to serve as a unifying account of brain functioning.

Appendices



References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function, 218*(3), 611-643.

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience, 30*(8), 2960-2966.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*(5), 789.

Bach, P., Gunter, T. C., Knoblich, G., Prinz, W., & Friederici, A. D. (2009). N400-like negativities in action perception reflect the activation of two components of an action representation. *Social Neuroscience, 4*(3), 212-232.

Bach, P., Knoblich, G., Gunter, T. C., Friederici, A. D., & Prinz, W. (2005). Action comprehension: deriving spatial and functional relations. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 465.

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass, 11*(7).

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695-711.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*(8), 706-712.

Berti, S., & Schröger, E. (2004). Distraction effects in vision: behavioral and event-related potential indices. *NeuroReport, 15*(4), 665-669.

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., & Rothwell, J. C. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Current Biology, 18*(10), 775-780.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*(5763), 1020-1022.

Blommaert, L., Coenders, M., & Van Tubergen, F. (2013). Discrimination of Arabic-named applicants in the Netherlands: An internet-based field experiment examining different phases in online recruitment procedures. *Social Forces, 92*(3), 957-982.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Current Biology, 17*(24), 2117-2121.

Campbell, M. E., & Cunnington, R. (2017). More than an imitation game: Top-down modulation of the human mirror system. *Neuroscience & Biobehavioral Reviews, 75*, 195-202.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping, 30*(8), 2313-2335.

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology, 17*(17), 1527-1531.

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1528), 2369-2380.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40*(1-2), 83-120.

Clark, A. (2013a). The many faces of precision. *Frontiers in Psychology, 4*, 270.

Clark, A. (2013b). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: the role of awareness. *Science, 280*(5360), 77-81.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: from origin to function. *Behavioral and Brain Sciences, 37*(2), 177-192.

Cooper, R. P., Cook, R., Dickinson, A., & Heyes, C. M. (2013). Associative (not Hebbian) learning and the mirror neuron system. *Neuroscience Letters, 540*, 28-36.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306-324.

Csibra, G. (2008). *Action mirroring and action understanding: An alternative account* (Y. R. P. Haggard, & M. Kawato Ed.). Oxford: Oxford University Press.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica, 124*(1), 60-78.

De Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology, 18*(6), 454-457.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *The Neuroscientist, 13*(6), 580-593.

Den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience, 30*(9), 3210-3219.

Den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2008). A dual role for prediction error in associative learning. *Cerebral Cortex, 19*(5), 1175-1185.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*(1), 176-180.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage, 55*(2), 705-712.

Doise, W., Csepeli, G., Dann, H. D., Gouge, C., Larsen, K., & Ostell, A. (1972). An experimental investigation into the formation of intergroup representations. *European Journal of Social Psychology, 2*(2), 202-204.

Dungan, J. A., Stepanovic, M., & Young, L. (2016). Theory of mind for processing unexpected events across contexts. *Social Cognitive and Affective Neuroscience, 11*(8), 1183-1192.

Everett, J. A., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience, 9*, 15.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Flannigan, N., Miles, L. K., Quadflieg, S., & Macrae, N. (2013). Seeing the unexpected: Counter-stereotypes are implicitly bad. *Social Cognition, 31*(6), 712-720.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*(1), 48.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*(2), 109-128.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815-836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1211-1221.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593-609.

Gazzola, V., & Keysers, C. (2008). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19*(6), 1239-1255.

Giudice, M. D., Manera, V., & Keysers, C. (2009). Programmed to learn? The ontogeny of mirror neurons. *Developmental Science, 12*(2), 350-363.

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*(8), 845.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Science, 25*(4), 565-610.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review, 111*(1), 3.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464.

Greven, I. M., Downing, P. E., & Ramsey, R. (2016). Linking person perception and person knowledge in the human brain. *Social Cognitive and Affective Neuroscience, 11*(4), 641-651.

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: an fMRI study. *NeuroImage, 21*(2), 744-750.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science, 35*(8), 1407-1455.

Hamilton, A. F. d. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience, 26*(4), 1133-1137.

Hastorf, A. H., & Cantril, H. (1954). They saw a game; a case study. *The Journal of Abnormal and Social Psychology, 49*(1), 129.

Heil, L., van Pelt, S., Kwisthout, J., van Rooij, I., & Bekkering, H. (2014). Higher-level processes in the formation and application of associations during action understanding. *Behavioral and Brain Sciences, 37*(2), 202-203.

A

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews, 34*(4), 575-583.

Heyes, C. M., & Ray, E. D. (2000). What is the significance of imitation in animals? *Advances in the Study of Behavior, 29*, 215-245.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): a study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*(8), 1555-1566.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108*(3), 687-701.

Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other's goals automatically biases the perception of their actions. *Cognition, 146*, 245-250.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3*(3), e79.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526-2528.

James, W. (1890). *The Principles of Psychology* (Vol. 2). New York: Henry Holt and Company.

Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology, 54*(5), 778.

Kerzel, D., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Launching the effect: Representations of causal movements are influenced by what they lead to. *The Quarterly Journal of Experimental Psychology, 53*(4), 1163-1185.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences, 11*(5), 194-196.

Keysers, C., & Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1644), 20130175.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Sciences, 8*(11), 501-507.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences, 15*(8), 352-357.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). The mirror-neuron system: a Bayesian perspective. *NeuroReport, 18*(6), 619-623.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*(3), 159-166.

Kilner, J. M., & Frith, C. D. (2008). Action observation: inferring intentions without mirror neurons. *Current Biology, 18*(1), R32-R33.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience, 29*(32), 10153-10159.

Kim, D.-Y. (2003). Voluntary controllability of the implicit association test (IAT). *Social Psychology Quarterly*, 83-96.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences, 27*(12), 712-719.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology, 26*(3), 371-376.

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron, 79*(5), 836-848.

Krall, S., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P., Eickhoff, S., Fink, G., & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function, 220*(2), 587-604.

Kwisthout, J., Bekkering, H., & Van Rooij, I. (2017). To be precise, the details don't matter: on predictive processing, precision, and level of detail of predictions. *Brain and Cognition, 112*, 84-91.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377-1388.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*(7), 1434-1448.

Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., & Fias, W. (2011). Inconsistencies in spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience, 7*(8), 937-950.

Mars, R. B., Sallet, J. m., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2011). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral Cortex, 22*(8), 1894-1903.

Martiny-Huenger, T., Gollwitzer, P. M., & Oettingen, G. (2014). Selective attention to in-and out-group members systematically influences intergroup bias. *Social Psychological and Personality Science, 5*(8), 936-943.

Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2012). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience, 8*(6), 623-631.

Miall, R. C. (2003). Connecting mirror neurons and forward models. *NeuroReport, 14*(17), 2135-2137.

Michotte, A. E. (1963). *The Perception of Causality* (T. R. Miles & E. Miles, Trans.). London: Methuen.

Mitchell, J. P. (2007). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex, 18*(2), 262-271.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews, 36*(1), 341-349.

Molenberghs, P., Halász, V., Mattingley, J. B., Vanman, E. J., & Cunnington, R. (2013). Seeing is believing: Neural mechanisms of action–perception are biased by team membership. *Human Brain Mapping, 34*(9), 2055-2068.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*(8), 750-756.

Nesdale, D., Maass, A., Griffiths, J., & Durkin, K. (2003). Effects of in-group and out-group ethnicity on children's attitudes towards members of the in-group and out-group. *British Journal of Developmental Psychology, 21*(2), 177-192.

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences, 110*(38), E3660-E3669.

Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition, 112*, 69-77.

Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology, 36*(1), 77-89.

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies *Progress in Brain Research* (Vol. 229, pp. 257-284): Elsevier.

Palmer, C. J., Seth, A. K., & Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. *Consciousness and Cognition, 36*, 376-389.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504-510.

Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., & Rowe, J. B. (2015). Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. *Journal of Neuroscience, 35*(25), 9255-9264.

Pineño, O., Denniston, J. C., Beckers, T., Matute, H., & Miller, R. R. (2005). Contrasting predictive and causal values of predictors and of causes. *Learning & Behavior, 33*(2), 184-196.

Pineño, O., & Miller, R. R. (2007). Comparing associative, statistical, and inferential reasoning accounts of human contingency learning. *Quarterly Journal of Experimental Psychology, 60*(3), 310-329.

Press, C., Heyes, C., & Kilner, J. M. (2011). Learning to understand others' actions. *Biology Letters, 7*(3), 457-460.

Quadflieg, S., Flannigan, N., Waiter, G. D., Rossion, B., Wig, G. S., Turk, D. J., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *NeuroImage, 57*(2), 549-557.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79.

Ratner, K. G., & Amodio, D. M. (2013). Seeing "us vs. them": Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology, 49*(2), 298-301.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory, 2*, 64-99.

Rivera, L. A. (2012). Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review, 77*(6), 999-1022.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131-141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*(9), 661.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*(4), 264.

Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology, 75*, 54-70.

Ryckman, R. M., Hammer, M., Kaczor, L. M., & Gold, J. A. (1990). Construction of a hypercompetitive attitude scale. *Journal of Personality Assessment, 55*(3-4), 630-639.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *NeuroImage, 19*(4), 1835-1842.

Saxe, R., & Powell, L. J. (2006). It’s the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science, 17*(8), 692-699.

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43*(1), 149.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2*, 395.

Shibata, H., Inui, T., & Ogawa, K. (2011). Understanding interpersonal action coordination: an fMRI study. *Experimental Brain Research, 211*(3-4), 569-579.

Shultz, T. R., Fisher, G. W., Pratt, C. C., & Rulf, S. (1986). Selection of causal rules. *Child Development*, 143-152.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605.

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *Journal of Neuroscience, 30*(9), 3531-3543.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage, 59*(3), 3050-3059.

Spunt, R. P., & Lieberman, M. D. (2013). The busy social brain: evidence for automaticity and control in the neural systems supporting social cognition and action understanding. *Psychological Science, 24*(1), 80-86.

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience, 23*(1), 63-74.

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1105.

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience, 15*(11), 745.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron, 59*(2), 336-347.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*(9), 1004.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*(5), 96-103.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149-178.

Turner, J. C. (1975). Social comparison and social identity: Some prospects for intergroup behaviour. *European Journal of Social Psychology, 5*(1), 1-34.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. *Progress in Social Psychology, 1*, 49-72.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences, 11*(4), 153-157.

Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin, 35*(3), 321-335.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychological Science, 19*(11), 1131-1139.

Van Bavel, J. J., Swencionis, J. K., O'Connor, R. C., & Cunningham, W. A. (2012). Motivated social memory: Belonging needs moderate the own-group bias in face recognition. *Journal of Experimental Social Psychology, 48*(3), 707-713.

Van Elk, M., Van Schie, H. T., & Bekkering, H. (2009). Action semantic knowledge about objects is supported by functional motor activation. *Journal of Experimental Psychology: Human Perception and Performance, 35*(4), 1118.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage, 48*(3), 564-584.

Van Pelt, S., Heil, L., Kwisthout, J., Ondobaka, S., Van Rooij, I., & Bekkering, H. (2016). Beta-and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience, 11*(6), 973-980.

Vander Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychological Science, 20*(6), 771-777.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences, 108*(51), 20754-20759.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 53.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598.

Westra, E. (2017). Stereotypes, theory of mind, and the action–prediction hierarchy. *Synthese*, 1-26.

Xiao, Y. J., Coppin, G., & Van Bavel, J. J. (2016). Perceiving the world through group-colored glasses: A Perceptual Model of Intergroup Relations. *Psychological Inquiry, 27*(4), 255-274.

Nederlandse samenvatting

In ons dagelijks leven zien we mensen om ons heen vaak allerlei soorten bewegingen maken. Stel je bijvoorbeeld eens voor dat je een vriendin ziet koken. Je ziet haar groenten snijden, een pan op het vuur zetten en door een saus roeren. In de tussentijd verwerk je al deze informatie: misschien vraag je je af wat je vriendin precies gaat maken, wat je zou kunnen doen om haar te helpen en of het gerecht lekker zal worden. Je hersenen komen met allerlei informatie om hierbij te helpen. Ze herinneren je eraan wat er de vorige keer gebeurde toen ze kookte en activeren kennis over bijvoorbeeld smaken en kooktechnieken. Daardoor begrijp je bijvoorbeeld meteen dat ze wil dat je een snufje van het zout waar ze naar wijst toevoegt aan de saus en verwacht je een heerlijk gerecht omdat je weet dat ze goed kan koken.

In een situatie als deze, maar ook in vele andere situaties, verwerken we informatie over het gedrag van anderen. Hoewel we dit meestal doen zonder dat we ons ervan bewust zijn, is deze vaardigheid cruciaal voor succesvolle sociale interactie. Het helpt ons niet alleen om te begrijpen wat er om ons heel gebeurt, maar ook om te voorspellen wat er daarna gaat gebeuren en hoe we ons eigen gedrag daaraan kunnen aanpassen. Door het gemak waarmee we dit doen lijkt het misschien eenvoudig, maar als onderzoekers begrijpen we nog relatief weinig van de sociaal-cognitieve processen die bij dit type verwerking betrokken zijn. Er moeten namelijk diverse soorten informatie geïntegreerd worden om te begrijpen waarom iemand een bepaalde handeling uitvoert en om zijn of haar volgende handeling te voorspellen. Het is dan bijvoorbeeld relevant om te weten wat die persoon daarvóór gedaan heeft, tot welke sociale groep hij of zij behoort en in welke context de handeling plaatsvond. Als we deze processen beter zouden begrijpen, dan zouden we niet alleen meer te weten komen over de werking van onze hersenen in het algemeen, maar ook over aandoeningen die ervoor zorgen dat mensen problemen ervaren met de verwerking van de bewegingen van anderen, zoals autisme of schizofrenie.

Daarom onderzocht ik in dit proefschrift of de snelheid waarmee we handelingen van anderen kunnen verwerken, en het neurale mechanisme dat hieraan ten grondslag ligt, uitgelegd kan worden vanuit de predictive processing theorie, die volgens diverse onderzoekers een algemene verklaring biedt voor de werking van de hersenen. Volgens de predictive processing theorie wordt informatie die binnenkomt via onze zintuigen steeds vergeleken met voorspellingen die ontstaan in cognitieve modellen. Als voorspellingen en binnenkomende informatie niet overeenkomen, ontstaan er voorspellingsfouten die in de hersenen opgelost moeten worden en dus extra verwerking vereisen. Het gevolg daarvan zou zijn dat de verwerking van verwachte gebeurtenissen sneller en efficiënter verloopt dan die van onverwachte gebeurtenissen.

De verschillende hoofdstukken in dit proefschrift beschrijven vier studies waarin we onderzochten of deze theorie inderdaad een verklaring kan bieden voor verschillende aspecten van de verwerking van de bewegingen van mensen om ons heen. Voor elke studie formuleerden en toetsten we hypothesen over de mogelijke eigenschappen van de cognitieve modellen die voorspellingen voortbrengen. We maakten gebruik van een experimenteel paradigma waarin proefpersonen keken naar animatiefilmpjes van mensen die bowlen. Op basis van verschillende soorten kennis, bijvoorbeeld over de spelers zelf, bouwden proefpersonen cognitieve modellen op die zorgden voor verwachtingen over de score die de spelers waarschijnlijk zouden behalen.

In **hoofdstuk 2** wordt een studie beschreven waarin we onderzochten of proefpersonen kennis over het vaardigheidsniveau van individuele spelers gebruiken bij het verwerken van uitkomsten van de bewegingen van die persoon. Ze zagen een ervaren of onervaren speler een hoge of lage score behalen en beantwoordden vragen over de speler of de score. We namen aan dat scores die onverwacht zijn op basis van kennis over de speler – zoals wanneer een onervaren speler een hoge score behaalt – zouden resulteren in voorspellingsfouten, die op hun beurt zouden leiden tot langere reactietijden. De resultaten toonden aan dat reacties op vragen volgend op onverwachte scores inderdaad langer op zich lieten wachten. Dit was specifiek het geval voor vragen over de speler, in overeenstemming met het idee dat deze informatie op een hoger niveau gerepresenteerd wordt dan informatie over de score. Een onverwachte score leidt volgens de theorie namelijk tot een voorspellingsfout die verwerkt wordt op een hoger niveau, in dit geval het niveau waarop informatie over de speler verwerkt wordt. Die verwerking zou er dan voor zorgen dat het langer duurt om informatie over de speler op te halen. Dit patroon zagen we niet in een vervolgexperiment waarin voorspellingen niet gebaseerd konden worden op kennis over de speler, maar op een gekleurd vak dat de kans op een hoge of lage score aangaf. We concludeerden dat een predictive processing model met een hiërarchische structuur en causale relaties tussen de niveaus in staat was het patroon van reactietijden in een set van drie experimenten te verklaren.

In de fMRI studie die beschreven wordt in **hoofdstuk 3** voerden proefpersonen dezelfde taak uit als in de vorige studie. We zagen dat onverwachte scores zorgden voor extra activiteit in de hersenen, die mogelijk nodig was voor het verwerken van voorspellingsfouten. Deze activiteit kwam voor een deel overeen met activiteit die we zagen tijdens een theory-of-mind taak. Dit suggereert dat voorspellingsfouten verwerkt worden in hersengebieden die vaak geassocieerd worden met sociaal-cognitieve verwerking en daarmee ook dat de predictive processing theorie mogelijk niet alleen het patroon van activatie verklaart in gebieden die betrokken zijn bij basale verwerking, maar ook in gebieden die betrokken zijn bij meer complexe, abstracte verwerking.

Hoofdstuk 4 beschrijft vervolgens een studie waarin we onderzochten of de cognitieve modellen die voorspellingen voortbrengen over de prestaties van anderen verschillen voor leden van onze eigen sociale group (de ‘ingroup’) en die van een andere sociale groep (de ‘outgroup’). Proefpersonen zagen leden van hun eigen en een ander team bowlen en voorspelden welke score elke speler zou behalen. De resultaten laten zien dat proefpersonen in staat waren scores van spelers van zowel hun eigen als het andere team te voorspellen, maar toch de prestaties van hun eigen team overschatten. Dat zou betekenen onze cognitieve modellen niet alleen informatie representeren over individuen, maar ook over de sociale groepen waartoe die individuen behoren. De resultaten suggeren daarnaast dat de overschatting van onze eigen sociale groep blijft bestaan wanneer we meer kennis opdoen over mensen uit andere sociale groepen.

In **hoofdstuk 5** wordt ten slotte een studie beschreven waarin onderzocht werd hoe kennis over verschillende soorten causale factoren in de omgeving gerepresenteerd worden in cognitieve modellen. In de bowlinganimaties die proefpersonen zagen waren steeds aanwijzingen te zien die de score voorspelden. Deze aanwijzingen waren gekoppeld aan factoren waarvan we aannamen dat ze wel (in het eerste experiment) of juist niet (in de volgende experimenten) gezien zouden kunnen worden als ondersteunende factoren in de causale keten van gebeurtenissen. In het eerste experiment was de aanwijzing bijvoorbeeld de kleur van de bowlingbaan, die gezien zou kunnen worden als een indicatie van de kwaliteit van de baan, terwijl de aanwijzing in het laatste experiment een gekleurd vak was dat los stond van de bowlingbaan en de speler. We verwachtten dat voorspellingen over een gebeurtenis gebaseerd worden op kennis over factoren die de gebeurtenis veroorzaken of op een andere manier een rol spelen is de causale keten. De resultaten suggereren echter dat geen van de situaties causaal geïnterpreteerd werd, maar dat proefpersonen toch in staat waren om de aanwijzingen te gebruiken bij het voorspellen van de score. Mogelijk betekent dit dat niet-causale factoren opgenomen kunnen worden in causale modellen, maar dat dit minder makkelijk gebeurt of ze dan minder voorspellende kracht hebben.

Deze studies samen ondersteunen het idee dat voorspellingen over de uitkomsten van handelingen van mensen om ons heen ontstaan in cognitieve modellen met een hiërarchische structuur en die kennis bevatten over individuen, hun sociale groep en de causale relaties tussen gebeurtenissen. Daarnaast suggereren ze dat uitkomsten van handelingen die onverwacht zijn op basis van de kennis die we hebben over de persoon die de handeling uitvoert, zorgen voor voorspellingsfouten die, in elk geval voor een deel, verwerkt worden in hersengebieden die geassocieerd worden met sociaal-cognitieve processen.

In het afsluitende hoofdstuk wordt besproken dat deze bevindingen in lijn zijn met andere theorieën over sociaal-cognitieve verwerking. Eén voorbeeld daarvan zijn associatieve theorieën, die ervan uitgaan dat associaties gevormd worden wanneer twee gebeurtenissen gelijktijdig plaatsvinden en vaak ook dat dit specifiek gebeurt wanneer er door dat geblijktijdig plaatsvinden een voorspellingsfout onstaat. Volgens sommigen wordt dit proces gedeeltelijk gestuurd door de kennis die mensen hebben over de causale structuur van de wereld: begrijpen waarom A leidt tot B, zorgt ervoor dat we B beter kunnen voorspellen op basis van A. Samen zouden deze theorieën een verklaring kunnen bieden voor de manier waarop cognitieve modellen opgebouwd worden.

Geconcludeerd wordt dat de predictive processing theorie een verklaring kan bieden voor de relatief abstacte social-cognitieve processen die betrokken zijn bij het verwerken van het gedrag van mensen om ons heen. Diverse aspecten moeten echter verder uitgewerkt worden, zodat de theorie beter vergeleken kan worden met andere theorieën. Het is nu bijvoorbeeld nog onbekend hoe abstracte informatie gerepresenteerd wordt in de hersenen en of de niveaus in de cognitieve modellen overeenkomen met specifieke hersengebieden. Als deze en andere aspecten verder uitgewerkt zijn en bepaald kan worden of predictive processing inderdaad de algemene werking van de hersenen zou kunnen verklaren.

Dankwoord

Het is af! Nu is het tijd om jullie te bedanken: de mensen die mij begeleid hebben tijdens mijn promotietraject, de mensen die me steunden als ik er zelf even niet meer zo in geloofde en de mensen die zorgden voor de nodige afleiding en gezelligheid.

In het bijzonder wil ik natuurlijk mijn (co)promotoren bedanken. **Harold**, dank voor al die ideeën die je tijdens onze bijeenkomsten wist te genereren en de verbanden die je legde waar ik ze zelf nog niet gezien had. Bedankt ook voor de prettige manier waarop jij de groep leidt en iedereen de vrijheid geeft om zijn of haar eigen ideeën te ontwikkelen. **Iris**, bedankt dat je mij met jouw open, kritische blik hebt geleerd mijn eigen gedachten en die van anderen te verhelderen. Ook bij mij komen mensen nu niet meer weg met een vage formulering, maar blijf ik doorvragen totdat ik weet wat ze precíes bedoelen. Bedankt ook voor de fijne gesprekken waarin je me soms het vertrouwen teruggaf als ik dat zelf even kwijtgeraakt was.

Bedankt aan de mensen met wie ik mocht samenwerken aan dit project. Als eerste natuurlijk mijn roomies Stan en Johan. **Johan**, bedankt voor alle discussies over de modellen die je uitwerkte en die ik door jouw geduldige uitleg beetje bij beetje beter ging begrijpen. Al die whiteboards die je voltekende zijn niet voor niets geweest! **Stan**, bedankt voor het delen van al jouw ideeën over de experimenten en over de wetenschap in het algemeen. Het was goed om zo nu en dan mijn twijfels daarover te kunnen bespreken. **Egbert**, bedankt dat ik altijd jouw hulp mocht inschakelen als ik weer eens vastliep met de fMRI analyses en voor de gezelligheid die je altijd bij je lijkt te hebben. **Sasha**, bedankt voor jouw hulp bij het opzetten van het fMRI experiment and **Olympia**, for continuing this project with me. **Birgit**, **Lukas** en **Lucas**, bedankt voor jullie hulp bij het testen van al die proefpersonen die vast niet altijd even gemotiveerd waren om in een donker lab honderden bowlingfilmpjes te bekijken. **Silvia**, thank you for setting up and running several experiments, for being the most enthusiastic student I supervised and for your sense of drama. I can still hear your 'Oh no, it's ALL wrong!' when I realise I made a mistake and somehow that always makes me smile.

I would also like to thank the other members of the **SENSe group**, who were part of the group before it was even called SENSe, members of the **BabyBRAIN group** and the **Computational Cognitive Science group**. **Sybrine**, **Denise**, **Lara**, **Peta**, **Irina**, **James**, **Simon**, **Marlene**, **Ricarda**, **Claire** and all the others: thank you for our interesting discussions and the fun we had inside and outside the DCC.

Wat ben ik blij dat ik mijn onderzoek heb mogen doen op het DCC, waar alles zo goed georganiseerd is. Daarvoor wil ik vooral **Jolanda**, **Vanessa** en **Maaike** bedanken. Daarnaast gaat mijn dank uit naar de mensen van de **Techical Support Group**, die altijd klaarstonden om problemen op te lossen. **Erik**, jij vooral ook bedankt voor het maken van al die bowlingfilmpjes die voor elk experiment toch nét weer een beetje anders moesten worden. **Miriam** en **Ronny**, bedankt voor de positieve en flexibele manier waarop jullie alles rondom de labs regelen.

Of course I would also like to thank my wonderful paranimphs. **Johanna**, jij was me altijd een stapje voor, waardoor ik bij jou kon afkijken hoe dat nu eigenlijk moest, zo'n promotieonderzoek. Samen tackelden we onderwijstaken en bureaucratie, sliepen we in een gebloemde caravan tussen de schapen en hadden we gezellige etentjes en koffiepauzes. Bedankt daarvoor! **Ezgi**, thank you for the long talks over coffee in which we discussed our struggles, science and the world in general. And thank you for all the fun we had filming all these 'kikker Kimmie' movies, thinking about our (still not-so-well-developed) plans for starting a company together, and on the Ratzeputz-filled trips to Berlin.

Dank ook aan de mensen die buiten mijn promotie om voor veel gezelligheid zorgden. **Ineke**, wat is het goed om te weten dat we altijd op elkaar kunnen terugvallen, zelfs als het er even niet van komt om af te spreken. **Simone**, bedankt voor jouw grenzeloze vertrouwen in mijn kunnen, hoe onterecht dan ook. **Tanya**, bedankt voor al die middagen en avonden dat we met een wijntje in de hand onszelf en de mensen om ons heen analyseerden. Ik ben blij dat ik jullie mijn vriendinnen mag noemen en hoop nog heel veel etentjes, feestjes en reizen met jullie mogen delen.

Datzelfde geldt natuurlijk voor jullie, **Rianne**, **Elly**, **Mirjam**, **Ellen** en **Britt**. Bedankt voor al die weekendjes, waarop we na al die jaren nog steeds zoveel te bespreken hebben dat ze eigenlijk altijd te kort zijn.

Bedankt ook aan mijn **familie** – de familie Heil, de familie Duijf en de familie Schiffelers, die ik binnenkort ook officieel mijn familie mag noemen – voor alle gezellige bijeenkomsten en jullie interesse in mijn onderzoek.

En aan mijn ouders, **Ben** en **Francien**, die mijn nieuwsgierigheid stimuleerden en die me leerden dat je nooit genoeg vragen kunt stellen (één van jullie nog net iets meer dan de ander). Ik kan me geen betere ouders voorstellen en het geeft me vertrouwen om te weten dat ik altijd op jullie kan terugvallen.

Bart, bedankt voor wie je bent. Bedankt voor al jouw steun. Bedankt dat ik mijn leven met jou mag delen.

Curriculum vitae

Lieke Heil was born on May 5th 1987 in Deurne, the Netherlands. In 2005, she started her bachelor in Psychology at Maastricht University. During this time, her interest in the relationship between brain functioning and human behaviour grew. After receiving her bachelor's degree (cum laude) in 2008, she was admitted to the Research Master in Cognitive and Clinical Neuroscience at the same university. As part of that master's programme, she went to Stockholm in 2010 for a research internship under the supervision of Torkel Klingberg at the Karolinska Institute. There she investigated working memory performance under stress in people dealing with burnout syndrome. She returned to Maastricht for a clinical internship at the Department of Medical Psychology at the Maastricht University Medical Centre and obtained her master's degree in 2011. Afterwards, she worked as a research assistant in Ivan Toni's group at the Donders Institute. In September 2012, she started her PhD project on action understanding in the context of predictive processing under the supervision of Harold Bekkering and Iris van Rooij, which resulted in this PhD thesis.

Publications

Heil, L., Hartstra, E., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (submitted). Processing of prediction errors in mentalizing areas.

Heil, L., Brecht, S., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (submitted). Biased person performance models: How ingroup overestimation persists in the face of outgroup individuation.

Kayhan, E., **Heil, L.**, Kwisthout, J., Van Rooij, I., Hunnius, S., & Bekkering, H. (submitted). Young children integrate prior probabilities and new observations to build predictive models of other's actions. Preprint available at https://doi.org/10.1101/365361.

Heil, L., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (2018). One wouldn't expect an expert bowler to hit only two pins: Hierarchical predictive processing of agent-caused events. *Quarterly Journal of Experimental Psychology*. Advanced online publication. doi:10.1177/1747021817752102.

Van Pelt, S., **Heil, L.**, Kwisthout, J., Ondobaka, S., Van Rooij, I., & Bekkering, H. (2016). Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience, 11*(6), 973-980.

Heil, L., van Pelt, S., Kwisthout, J., van Rooij, I., & Bekkering, H. (2014). Higher-level processes in the formation and application of associations during action understanding. *Behavioral and Brain Sciences, 37*(2), 202-203.

Ferraye, M. U., Debû, B., **Heil, L.**, Carpenter, M., Bloem, B. R., & Toni, I. (2014). Using motor imagery to study the neural substrates of dynamic balance. *PloS ONE*, *9*(3), e91183.

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:
http://www.ru.nl/donders/graduate-school/phd/

DONDERS
INSTITUTE

ISBN 978-94-6284-156-7

Radboud University Radboudumc